

nature

21/28 December 1978

No end in sight to jobs gloom

FOUR years ago (Vol. 253, p. 79) we looked at the fluctuations in job opportunities in British universities since 1960. Now is a suitable time to bring the record up to date. The graph shows in a schematic way numbers of tenured posts available up until the end of 1978.

The technique used is the same as that employed in 1975. The classified advertisements in *Nature* for two weeks, in May and November, of each year were examined. Posts at or above the level of lecturer at a British university scored one. Temporary posts scored half. The figures for May and November were added together. Various objections can be raised; that not all jobs will be advertised in *Nature*; that the method of measurement, though consistent, is crude; that many universities have research-council establishments attached to them which perhaps should also be included; that there are large research-council and other establishments not attached to universities but which offer at least a semi-academic environment. Nonetheless, it is likely that the graph is a moderately good guide to the ups and downs of employment prospects for the academically inclined.

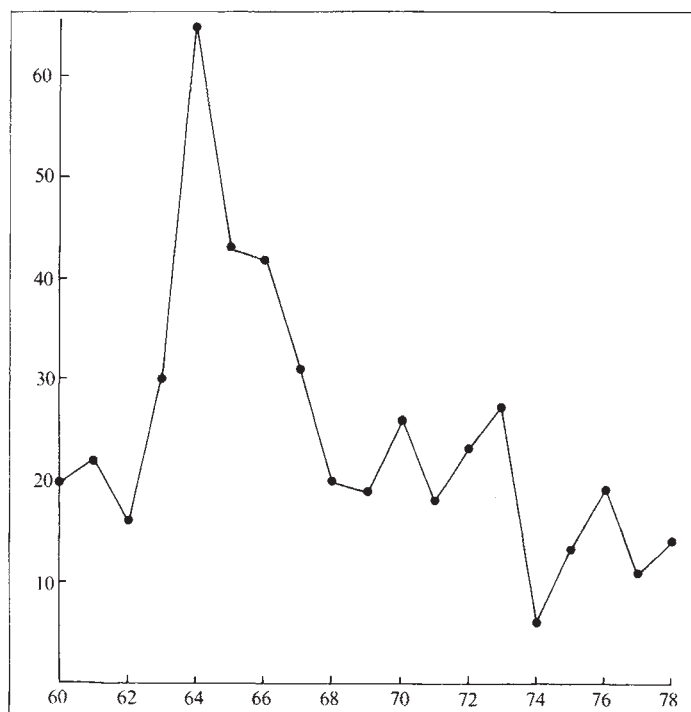
The enormous spike in the mid 1960s need not detain us long. This was the time when new universities were rapidly coming on stream and recruiting with immense enthusiasm. This cohort of new-university staff is now largely in its 30s and 40s and unlikely to retire for some time. In the late 1960s recruiting seemed to return to its old level but then in 1974 many university posts were 'frozen', owing to a major economic crisis.

What has happened since then? There has been no obvious rebound; it must be assumed that many of the frozen posts have never re-appeared. In fact the level of recruitment in the past four years looks unmistakably down on that of the early 1970s, for which the present low-level of retirements from universities must

take part of the responsibility. Nor is there much comfort just over the horizon. The numbers of students choosing to do science at university is certainly not growing quickly enough to warrant major increases in staff; indeed in recent years numbers have often declined. And the demographic signs are unmistakably that young people of university age will be substantially less in number in ten years' time than they are now, so teaching opportunities are hardly likely to increase.

At present a large and growing number of post-doctoral research workers find themselves caught in the trap. Tenured posts are few and far between whilst short-term contract research merely delays the day of decision and offer no security. The recently formed Association of Researchers in Medical Sciences, intended to articulate the concern of medical researchers thus trapped, could easily have a counterpart in other branches of science.

No one simple solution will emerge to this very serious problem. The encouragement of early retirement is likely to increase, and will help, at least temporarily. Specific initiatives have helped in some particularly difficult circumstances. But in the long run the responsibility may lie as much as anywhere on those tenured university staff who take on new graduates to study for a Ph.D. In the flush of satisfaction at having found just the right person to solve the problem in hand, it is all too easy not to make discouraging noises at the same time about long-term job prospects. Research students should be told in no uncertain terms that a Ph.D. is no longer a guarantee of entry into the academic world and in some circumstances over-qualifies for non-academic jobs and so could even be a liability. Any research supervisor who fails to make potential students think twice about the path on which they are entering is failing in one of his most important responsibilities. □





US extends recombinant DNA controls to private industry

A SERIES of moves was announced this week by Mr Joseph A. Califano, Jr, US Secretary of Health, Education and Welfare, designed to extend the guidelines currently covering federally-sponsored research involving the use of recombinant DNA techniques to all such research carried out in the private sector.

In particular, Mr Califano has directed the Food and Drug Administration to use its statutory authority to require that all recombinant DNA research submitted to satisfy regulatory requirements—for example in order to licence a new drug—comply with the guidelines that have been developed by the National Institutes of Health.

And as regards research in the private sector which would not be submitted to the FDA in this way, he has requested the Environmental Protection Agency to take all the action it can "to require that recombinant DNA research conducted by the private sector complies with the NIH guidelines."

Mr Califano also announced that he has approved final guidelines prepared by Dr Donald Fredrickson, director of NIH, that significantly revise the safety requirements for conducting recombinant DNA research.

As expected, the revised guidelines substantially broaden public representation on the Recombinant DNA Advisory Committee, which assists the NIH in administering the guidelines. It also increases public participation in the administration of the guidelines by local biosafety committees, as well as public access to information about recombinant DNA research.

Finally, since the revised guidelines require the director of NIH to determine that any proposed action under the guidelines "presents no significant risk to health or to the environment," Mr Califano has directed the NIH to formulate a balanced programme of risk-assessment experiments at NIH directly, or under NIH support.

"In my view, the more risk assessment experiments NIH carries out, the better we will be able to judge whether the guidelines—and actions taken under them—afford appropriate protection for health and the environment," he says in a memorandum to Dr Fredrickson and Dr Julius Richmond, Assistant Secretary for Health.

In a public statement announcing these various actions, Mr Califano says that the final version of the NIH guidelines, based largely on proposed

revisions published in July, relax in two major respects the guidelines that were placed in effect in 1976.

The first is that the revisions exempt altogether five categories of experiments, which the NIH has concluded present no health risk. This will exempt approximately one-third of the research now covered by guidelines.

The second relaxation is a general easing of restrictions on other permissible experiments. Almost all categories of research are assigned to physical containment and/or biological containment levels at least one step lower than in the 1976 guidelines.

"Since the likelihood of harm now appears more remote than was once anticipated, the scientific community has now concluded that this downgrading is appropriate," Mr Califano says.

A number of significant changes to the proposed revisions have been introduced to the ways in which the guidelines are administered, following a review carried out by a departmental committee chaired by Mr Peter Libassi, DHEW's general counsel.

In particular, whereas the 1976 guidelines had no specific requirements for public participation, it is now specified that 20% of the members of local institutional biosafety committees (IBCs) must represent the public, and have no connection to the institution.

It is also now required that the same proportion of the Recombinant DNA Advisory Committee be "persons knowledgeable about such matters as applicable law, standards of professional conduct and practice, public and occupational health, and environmental safety". And it is recommended that one member be a "non-doctoral" person from a laboratory staff.

To take in these extra interests, the membership of the committee is to be increased to 20, and the new members are expected to be announced by Mr Califano in the near future. Furthermore all federal agencies represented on the Federal Interagency Advisory Committee will have non-voting members on the RAC.

One result of demands for increased public participation is that all important records relevant to research carried out under the guidelines must be made public. In particular, the bulk of IBC records must be made available to the public, and problems, violations of the guidelines, illnesses and accidents must be reported to NIH.

Major actions under the guidelines—such as decisions to approve on a case-



Califano: more risk assessment experiments

by-case basis experiments that are generally prohibited, to exempt additional categories of research from the guidelines, or to permit the insertion of genes into new types of bacteria—cannot be taken without the advice of the RAC, with public and federal agency comment on the proposals.

In addition, the Director of NIH cannot approve any proposed actions unless he determines "that they present no significant risk to health or the environment". This implies the need for some type of risk assessment.

In a statement accompanying the final version of the guidelines, Dr Fredrickson says that one reason for distinguishing between the hazards of different organisms on the basis of phylogenetic relatedness—a subject recently debated in Britain's Genetic Manipulation Advisory Group (GMAG)—is that the more closely related the species, the more likely that polypeptide hormones or related proteins would be pharmacologically active, particularly in the case of recombinant DNA integrating into the human genome.

"This matter lies in a crucial area of risk analysis and will undoubtedly continue to be a subject of both further debate and improved understanding in the coming months", Dr Fredrickson writes.

The final guidelines also include detailed presentation of the types of responsibility to be delegated to the local biosafety committees. Dr Fredrickson says that although this will place a burden on the human and financial resources of the institution, the responsibility is better delegated than retained at the federal level.

A new proposal will allow NIH to require prior approval of all recombinant DNA research if the institution fails to comply with the guidelines.

David Dickson

UK researchers claim rights over short-term contracts

At least two scientists believe universities and research councils are treating post-docs shabbily, writes **Joe Schwartz**

Two British scientists are turning to the courts and industrial tribunals to find redress over dismissal at the end of short-term research contracts. Dr Desmond Turner has brought a tribunal hearing against the University of Sussex; and Dr Amiram Ur is bringing another, continuing this week, against the MRC's Clinical Research Centre in Harrow.

Dr Turner, 39, was asked to leave the University of Sussex in June at the end of an MRC-funded project to isolate gastro-intestinal hormones involved in the regulation of insulin secretion. Dr Ur, 44, was also asked to leave, in his case after limited-term contracts totalling eight years' work with the MRC at Harrow.

Both have been productive scientists. Turner has published 15 papers relating to the hormonal regulation of insulin secretion over a period of 14 years on short-term contracts. Ur, a bioengineer, has numerous patents, most notably an impedance measuring device for monitoring blood coagulation and bacterial cell growth which is patented in 18 countries and is marketed in the UK under the names Biobridge and Bactbridge.

Turner and Ur have found themselves at the centre of a legal and political thicket. First, the legal status of short-term contract workers is not at all clear. Despite the fact that post-doctoral fellowships, for example, extend only for an agreed fixed term, there appear to be rights under the Employment Protection Act, 1975, to redundancy pay and claims over unfair dismissal at the end of the contract. But increasingly employers—universities and research councils—have been introducing a 'waiver clause' into the contracts which nominally eliminate those rights.

Turner says: "There is no justification for the use of special forms of contract that deprive employees of their normal rights under the law. . . . We need to bring to public notice the shabby way in which institutions are treating temporary appointments".

On the one hand it would seem sensible that the granting institutions should be free to allocate money according to their own perception of scientific merit without fear of redress. On the other, it would also seem fair that the increasing number of experienced post-doctoral and other short-term contract researchers (perhaps 10,000 in the UK) should find some

security of employment. Mr P. C. Stephenson, Secretary of Science at the University of Sussex and responsible for Turner's employment, recognises the problem but feels that the Government and the Association of University Teachers (AUT) are responsible for creating a coherent national policy. "We are aware of the problem but we need to find the right place to start the argument," says Stephenson.

The legal position of contract workers became more complicated on 26 July 1977, when an Industrial Appeals Tribunal ruled in *L. D. Dixon v. BBC* that Dixon's contract was not fixed term during its tenure, thereby invalidating the waiver clause, but that it became fixed term at its point of expiration, thereby entitling Dixon to redundancy pay. In October of this year, an appeal court overruled the Appeals Tribunal and decreed that the waiver clause as customarily employed was valid.

Following this decision the AUT withdrew from Turner's case. However, Turner, with the aid of Brian Salter, chairman of the research sub-committee of the AUT's Sussex University branch, took his case to the tribunal hearing on 27 November, 1978. "We felt we still could make a case and we wanted the tribunal to commit itself to say that the appeal court ruling on *Dixon v. BBC* was relevant to research workers," says Turner. The tribunal has not yet made its decision.

Salter, who is encouraging researchers to press redundancy claims, agrees that a coherent national policy is called for, but says that the AUT is not sufficiently aware of the conditions of research workers. Salter feels that bridging loans between contracts, re-organisation of the nation's research into core areas located within universities and a well defined career structure alongside the normal university lecturer scales is needed. "The Research Councils will also have to accept that people have to be hired even when they get older", says Salter.

Salter and Turner both think that a closed union shop may be the only way for short term contract workers to get any control over their jobs at all. "Under conditions where there are so many out of work scientists, the jobs will go to the cheapest bidder to the detriment of us all", says Turner.

The case of Dr Amiram Ur, being

heard again last week in London by an industrial tribunal after two months adjournment, also raises fundamental questions. In July 1974, the MRC introduced 'Establishment Code 335' which radically revised the career structure for non-clinical staff. Citing the greatly reduced university job opportunities as a reason, the MRC instituted a procedure under which, after four years, "awards of unlimited tenure will be made on average to 75% (taking one year with another) of the number of staff awarded limited-term appointments."

Approximately 600 people were affected by the order, which would result in 450 tenured posts but also redundancy for 150 scientists. Changes in the Establishment Code are negotiated agreements and as the negotiating body, the AUT is implicated in this ruling. But as one experienced observer noted, "This AUT negotiated tenure procedure is really a dismissal procedure and doesn't meet the minimum legal requirements of a dismissal system." Ur was one of the MRC's employees that came under the new ruling.

On 29 March, 1974, Ur was transferred from clinical to non-clinical status which brought him under the new requirements imposed later in the year. In November 1976 Ur's mandatory application for tenure was denied and under the new terms of service was given a year to find another job. Ur protested the decision citing deficiencies in MRC procedure beginning with his transfer to non-clinical status. This failing, Ur was dismissed in November 1977. The first tribunal hearing on the case was scheduled for April 1978. After an MRC request for postponement, in June 1978 the MRC offered Ur a settlement, which included back pay, reinstatement on a three-year contract, legal costs up to £4,000, return of pension rights and an elaborate and costly procedure to "investigate the grievances for and against Dr Ur".

The grievance procedure provided for a three-person committee, one member selected by Ur, one by the MRC and a chair person by mutual agreement. In case of difficulty with the latter choice, nominations were to be received from the President of the Royal Society. The committee was to have a paid secretary and to have had the ability to call and pay witnesses. However, its findings were to be confidential to the MRC and council actions based on the report were to be final and binding. Ur refused the settlement.

Meanwhile, the hearings of Ur's tribunal are expected to carry over until March. □

Better Med than Dead

Vera Rich reports on Israel's plan to save the Dead Sea from drying up by building a canal linking it to the Mediterranean

ENVIRONMENTAL protection is normally one of the first victims of a war economy. It would be tempting, therefore, to interpret the Knesset motion tabled last month by former Foreign Minister Yigal Allon, to avert ecological disaster by constructing a canal linking the Mediterranean and the Dead Sea, as a sign of approaching political normalisation. Ironically, this is by no means the case; indeed, according to some Israeli sources, the signing of a peace treaty with Egypt would effectively block any such plans.

The idea of such a canal is not new. It was first proposed at the end of the last century by Theodore Herzl, the Zionist leader who first envisaged a future Jewish state. Since then, a number of plans have been proposed, including a grandiose scheme for a ship-canal down from Haifa, via the Sea of Galilee, lower Jordan and the Dead Sea, down to the gulf of Aqaba. Present plans, however, and there are five under consideration by the Israeli government, with three possible routes, concentrate rather on the possibility of using such a canal as a source of hydro-electric power which could produce up to 10% of Israel's present generating capacity.

The Dead Sea lies some 400m below the level of the Mediterranean. Theoretically, therefore, such a canal could descend via a number of hydro-electric cascades. Such a project is highly favoured by Israel's anti-pollution lobby, notably by Dr Anthony Pirano of the Techion, who has worked out his own "integrated" scheme, with desalination of the water at the inlet end, and use of the proposed storage reservoir at Arad as a winter spa. Other proposals are somewhat less elaborate, but all would admit that the construction of such a canal would not be at present justifiable economically, on the basis of the energy produced alone.

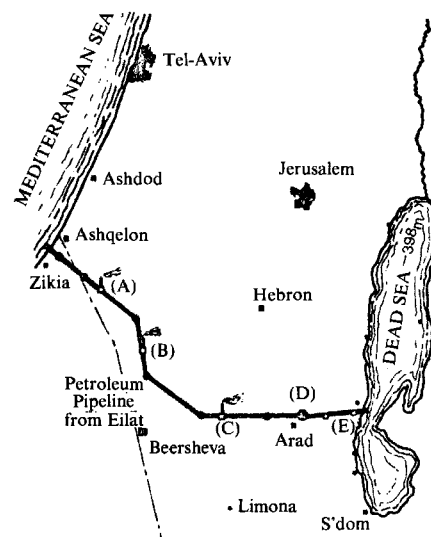
Replying in the Knesset to Mr Allon's motion, the Energy Minister, Yitzhak Moda'i, stressed this point. Electricity, he said, can still be generated more cheaply by conventional power-stations using fossil fuels. The main purpose of the canal would not, however, be the hydroelectric scheme, although, as Mr Joseph Vardi, official spokesman for Israeli Ministry of Energy and Infrastructure told *Nature*, power production would be a spin-off which would make the canal economically feasible. Its prime purpose would be to recharge the Dead Sea.

Of recent years, Israel's use of the Sea of Galilee as the water source of the whole country has led to a significant drop in the level of the lower Jordan and a consequent fall in the level of the Dead Sea. The arid climate of the Dead Sea region causes a high evaporation rate—a fact utilised by the Dead Sea potash plant at S'dom, which is based on the natural evaporation of brine from shallow pans. Already the receding shore-line has left the S'dom works some distance "inland," and it has been necessary to construct a new pumping station and canal to bring the brine to the evaporation pans.

Although the Dead Sea Works has thus solved their own immediate problem, this is no remedy for the ecological changes produced by the recession of the Sea. According to Mr Allon, the situation is becoming "urgent" due to continuing progress by the Jordanians in their own water projects. The diversion of the Yarmuk river (a tributary of the lower Jordan) via the East Ghor canal for irrigation purposes is almost complete. Half the streams flowing into the Jordan from the east have already been dammed, and the damming of the remainder is proceeding at full speed. Once these Jordanian schemes are complete, said Mr Allon, the lower Jordan will disappear completely, save as a drainage channel, the level of the Dead Sea will fall completely and the southern part of the present area of the Sea would become a salt desert. This, as Mr Moda'i admitted in his reply, could cause "tremendous" ecological damage, which could not be easily repaired.

Use of the Jordan water by Israel's eastern neighbours has always been a sensitive issue. The "war over the waters" between Israel and Syria, was one of the prime causes of tension which led ultimately, to the Arab-Israeli conflict of 1956. Since it is not the loss of the lower Jordan itself, but the drying up of the Dead Sea that is seen as the ecological hazard, the construction of the proposed canal would clearly remove at least one potential cause of future conflict between Israel and Jordan.

A speedy separate peace with Egypt, however, would almost certainly force any such construction project to be deferred. Not for any political reason, but simply because Israel's resources, both human and financial are limited, indeed, over-extended. Military withdrawal from the Sinai would entail the construction of



Dr Pirano's proposed canal route

new defence installations in the Negev.

Furthermore, troops stationed in the Sinai have left their families at home, generally in the populous areas in the north. If transferred to the Negev, however, they would probably bring their families south, with consequent development of such towns as Be'er Sheva and Dimona, as well as less fortunate settlements such as Yeruham which have never really become fully viable. Under such circumstances, Ben Gurion's vision of the Negev as the heartland of Israel's prosperity could well come true—but at the cost of resources which could be otherwise used for the proposed canal.

According to some Israeli sources, there is considerable pressure "from outside" for a post-treaty Israel to develop a joint nuclear power capability with Egypt, the stations being situated in the Sinai, as a further incentive to peace. Such a project running into "billions of dollars" would be of considerable interest to foreign investors. Israeli attitudes to nuclear power are generally cautious—particularly since the proposed coastal site for one nuclear station between Ashqelon and Ashdod (see map) was found to be earthquake-prone. Until now, however, costs have been too high for Israel to consider any major nuclear power programme.

Should the proposed joint Israel-Egypt Sinai scheme become a reality, the proposed canal would be less important from the energy point of view, and with still more of Israel's construction workforce committed to the Sinai power-stations, its early construction would seem most unlikely.

The ecological threat to the Dead Sea would remain. It would be ironical if the peace-makers who urge Israeli-Egyptian nuclear cooperation should prove later to have laid the foundations for a second "war of the waters". □

Inflation forces new tactics on US environmentalists

LAST SUMMER the environmental sciences board of the US National Academy of Sciences prepared a list of topics requiring its attention. At the top of the list was the potential threat to schemes for environmental management and regulation posed by pressures on the economy, and the political measures which might arise.

The board's concern reflected a debate that has been simmering in Washington since President Carter's administration began its attempts earlier this year to take a "firm grip" on the problem of inflation. A major component of the ensuing debate has been whether environmental controls are inflationary; and if so, what should be done about them.

Concern among environmentalists erupted briefly at the beginning of the autumn when Mr Robert Strauss, then the administration's chief inflation fighter, let slip his view that environmental regulation was one of the first areas in which cut-backs could be made.

Since then the rhetoric has been softened. Administration economists now talk about the need to demonstrate the cost-effectiveness of proposed regulations; and both regulators and environmentalists have won a number of important battles. But they are now on the defensive.

On balance, 1978 has not been a bad year for environmentalist causes. Though backing away from some commitments, President Carter has put up a strong fight against many potentially damaging water projects opposed by groups such as the Sierra Club, and recently signed into law the preservation of large areas of Alaska.

Furthermore last week the Tennessee Valley Authority announced after years of opposition, that it was prepared to invest over \$1 billion to clean up air pollution at ten of its power plants, a settlement described by the Environmental Protection Agency as "the largest ever made with a major source of air pollution".

However just as four years ago environmental regulations were being blamed for the problems of increasing unemployment, so the same is now being widely said about their contribution to inflation; industry claims, for example, that meeting new controls means taking away money from more productive areas of investment.

In reply, environmentalists claim not only that such regulations make a very small contribution to the increase in the consumer price index—estimated at less than 0.5%—but that if factors such

as reduced health care costs are taken into account, the benefits of pollution control measures outweigh the costs.

"Most important, I believe that the public wants those benefits and is prepared to pay for them," says Douglas M. Costle, administrator of EPA. A recent survey conducted by Resources for the Future found that 62% of those questioned were prepared to pay higher prices to reduce pollution.

But to answer the charges being laid against the regulators, convictions previously expressed in qualitative terms are now having to be quantified. "We need to focus more on the adequacy of data, their interpretation, the need for risk-benefit analysis from a systems view," says Dr Gilbert S. Omenn, assistant director of the President's Office of Science and Technology Policy.

This need is reflected in the shifting strategies of environmentalist groups. Initially reluctant to place quantifiable values on that which they are trying to defend, this is now becoming necessary to respond to proposed "anti-

'We are going to insist on better data on the cost-effectiveness of environmental control measures. I do not see this as necessarily bad, unless the pendulum goes too far'

—Prof John Cantlon

inflation" measures. The Natural Resources Defense Council, for example, one of the strongest such groups, is setting up a working group on the economics of the environment, and others are moving in this direction.

A major problem, however, is that whereas the likely costs of any new regulation are relatively easy to predict, the same is not true for the other side of the equation. "There is no question that trying to get at the benefits side of the picture is a very much tougher job of analysis and study," says Professor John E. Cantlon, professor of botany at Michigan State University, and chairman of the NAS's environmental sciences board.

"However it is something that needs to be done. The scientific community should not stand around wringing its hands, but should start getting research groups interested in the problem," he says.

"The climate in government and in industry broadly—and I would even say in society—is that we are going to insist

on better data on the cost-effectiveness of environmental control measures. I do not see this as necessarily bad, unless the pendulum goes so far as to call into question genuine concerns about internalising the social costs of production."

Others are concerned at this happening. Speaking at a meeting last week organised by a group called Environmentalists for Full Employment, Ms Peggy Seminario, an industrial hygienist with the American Federation of Labour and Congress of Industrial Organisations (AFL-CIO), claimed that many companies were using inflation as an excuse to resist further regulation, often using suspect data to support their arguments.

"At the heart of the debates on cost estimates, there are not many numbers, and the numbers are not that good. In 1974 environmental impact statements claimed the cost of meeting proposed vinyl chloride standards would be between \$65 million and \$90 million; recently, after the standards were introduced, it has been estimated that the costs to industry of meeting the regulations was less than \$1/2 million."

But the increasing incursion of economics into the environmental debate is also raising difficult issues within the environmentalist movement. Some have argued, for example, that the price of energy should be allowed to rise so that solar energy, already environmentally preferable to more conventional sources, can become economically competitive as well.

Others dispute this strategy, claiming that any increase in energy prices has an unfair impact on the poor. "The net result of present policies is to force people at the low end of the income distribution to lose food to pay for solar energy," Dr Gar Alperovitz, co-director of the Exploratory Project for Economic Alternatives, told the EFFE meeting.

"It is time for the environmental movement to take responsibility for what it is saying. We must reconsider the relation between strategies on prices and the desirable goals of conservation; for example there are other ways of getting to solar energy than by supporting higher prices."

Whichever way the various debates go, they seem destined, in Washington at least, to be increasingly fought on a terrain of economic argument which few can afford to ignore. As one Congressional aide put it last week: "Economists now seem to be playing the same role in Washington that scientists used to play in the 1960s."

David Dickson

news in brief

● **Death halts Greenpeace action over nuclear waste:** The first ship to carry nuclear waste from Japan to the reprocessing plant at Cap de la Hague near Cherbourg in France, has had to alter course for Barrow-in-Furness in England after one of its senior technical experts, Mr Robin Dale Hartley, died. Preliminary reports suggest that he died of a heart attack while in his bunk. Mr Hartley was an electrical engineer responsible for maintaining the cooling systems of the 26 containers of fuel rods of uranium oxide aboard the Pacific Fisher. British Nuclear Fuels Ltd, which had chartered the vessel, said that it expected that an inquest into the death might be necessary.

The consignment of nuclear waste is to be reprocessed at Windscale in Cumbria and at Cap de la Hague, under the terms of a contract between Britain, France and Japan by which each of the two European countries is to reprocess 1,600 tons of Japanese uranium oxide over a nine-year period from 1982. Although Britain has been taking nuclear waste from Japan since 1969, this is the first consignment to go to France.

The Greenpeace conservation group had planned to harass the Pacific Fisher as it entered Cherbourg, as part of a demonstration against shipping nuclear waste to France which involved 18 different French anti-nuclear and other political organisations. They were going to use their converted trawler the Rainbow Warrior which successfully frustrated the seal cull in the Orkneys last October.

Mr David McTaggart, European director of Greenpeace, said they were planning to run inflatable dinghies in front of the Pacific Fisher. Even though the ship is not going straight to Cherbourg, Mr Peter Wilkinson, Greenpeace director in charge of the anti-nuclear campaign, said they were still going ahead with a demonstration with the French organisations, but they were not going to demonstrate at Barrow-in-Furness, in deference to the dead man's family.

● **Museum can show Darwinian evolution:** The Smithsonian Institution's National Museum of Natural History is not required to depict the Biblical version of the creation of the universe as part of a scientific exhibition on evolution, a court in Washington ruled last week. Two religious groups claimed that a \$1½ million exhibition, which opens next May, violated the Government's role of religious neutrality by presenting Darwinian evolution as "the only credible theory of the origin of life". They asked the court either to ban the exhibition, or to require the museum to spend a comparable amount of money illustrating the Biblical account of the origins of man.

In rejecting both claims, US District Court Judge Barrington D. Parker said that the primary effect of the evolution exhibits was not to advance a religious theory, or to inhibit plaintiffs in their religious beliefs. "Even accepting their argument that evolution is hostile to their beliefs

as to creation, this impact is at most incidental to the primary effect of presenting a body of scientific knowledge."

● **Thumbs up for re-usable satellite:** A favourable report on Britain's involvement in the MRS re-usable satellite project is now being prepared by a Science Research Council team and is to be presented to the council's astronomy, space and radio board in the new year. The Multi-discipline Refurbishable Satellite project, which could allow different instrument packages to be replaced while a spacecraft is in orbit, makes use of the Multi-Mission Satellite (MMS) being studied by the Americans. If it is given the go-ahead by NASA, this would lead to the setting-up of a joint UK-US feasibility study that could see the development of the revolutionary new satellite at a cost of several tens of millions of pounds.

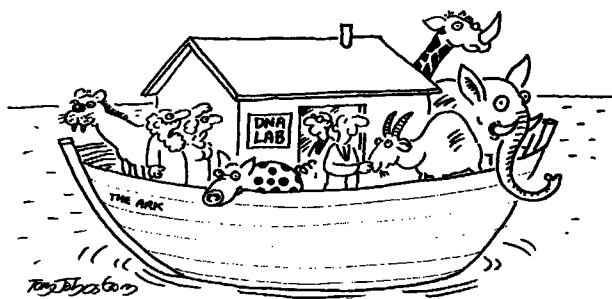
Recently a team at the SRC's Appleton Laboratory, in collaboration with British Aerospace engineers at Stevenage, has been considering the most testing shapes and types of various scientific instrument packages that would have to be fitted to the craft. And it is now known that no serious drawbacks have been found to the various tricky configurations considered and a favourable report is to be presented to an informal meeting of the ASR board on 10 January. A more detailed report is likely to be presented to a future formal meeting of the board.

● **First evidence of gravity waves:** A team of astronomers at the University of Massachusetts at Amherst claims to have discovered evidence of gravity waves, whose existence was first predicted in Einstein's general theory of relativity 60 years ago, but has yet to be demonstrated. Using the 1,000-foot diameter radio telescope at Arecibo, Puerto Rico, Dr Joseph Taylor, Professor of Astronomy at Massachusetts, Dr Peter McCulloch, on temporary leave from the University of Tasmania, and Lee Fowler, a graduate student, monitored the radio emissions from a pulsar, discovered in 1974, which is orbiting another massive object. According to the general theory of relativity, the orbiting pulsar should radiate gravitational energy which would cause the two bodies to move closer together and thus increase the speed of rotation, which is currently a cycle every eight hours. Relativity predicts that the period of rotation should decrease by one-tenth of a millisecond each year. The astronomers found that the pulsar has slowed down by four-tenths of a millisecond, $\pm$ one tenth, in the four years since its discovery.

The results were presented at the symposium on relativistic astrophysics at the Max Planck Gesellschaft Astrophysik in Munich this week. The research workers say that they have been able to rule out other factors, such as tidal movements on the two bodies, which might have caused the changes. Calculations of the periodicity of the pulsar, which circulates its companion at 0.1% the velocity of light, have been complicated by the fact that its orbit processes at a rate of 4 degrees a year.

One physicist pointed out last week that a rotating pulsar was an almost perfect system by which Einstein's theory could be tested, since his theoretical framework was based on the concept of observers with clocks—in this case the rotating neutron star—falling freely through space.

● **Correction:** In the article "Czech Chartists claim two died in nuclear accident" (*Nature*, 276, 7 December, 1978, p.551), the fifth paragraph should commence, "A year and six weeks later, however, disaster struck again (according to the Chartists)".



"Maybe it wasn't such a good idea to take along the two scientists!"

Science in China: is 'the bumpy road' better?

JOSEPH NEEDHAM'S articles on science in the People's Republic of China (31 August, 1978 page 832) are for the most part informative discussions of recent changes in the organisation and process of Chinese science. However, Dr Needham's report does not accurately reflect important aspects of the political-social history of the past few years in China, and as a result draws a somewhat unbalanced picture of the evolving role of science. This is largely due to Dr Needham's uncritical acceptance of the position of the current leaders with regard to the role attributed to the "Gang of Four" (G₄), especially with respect to their influences on science and technology.

A delegation of 12 members of Science for the People visited China from 4 June to 2 July at the invitation of the Scientific and Technical Association of the PRC to study scientific and technical decision-making in agriculture and food production. We travelled from Hainan Island in the south to as far north as Peking, visiting universities, research institutes, People's Communes, and factories and interviewing researchers, peasants, workers, and local and national officials. We saw much evidence of the Chinese leadership's decision to place great emphasis on science and technology to accelerate modernisation. We would like to analyse some of the important points made by Needham and present a somewhat different view of the way science is developing in China.

Needham essentially repeats the current official line in China that the G₄ caused the economy to stagnate, that their ideas were heretical to all Maoist thought, and that they nearly brought science and technology to a standstill. For instance, he agrees with the assertion that the G₄ "brought the economy to the brink of economic collapse". This simply was not so. Between 1965, the last year before the Cultural Revolution, and the most recent years for which we have estimates, production of electricity increased 157% (1974), production of steel 162% (1974), coal 76% (1974), oil 619% (1975)², and machinery 350% (1975)³. Overall industrial production was 7% higher during the first half of 1976 than in the same period of 1975⁴.

Needham asserts that "the cutting off of funds for all higher institutions was, under the G₄, a matter of course", and "necessarily in alliance with this was a general despising of science and scientists". We obtained quite a different picture. Needham says the G₄ exercised a dominant influence in China during the years 1974-6. We received

China is bent on a great leap forward in science and technology—but it should not be at the expense of its peasants or workers



By Stephen Risch, Michael Hansen, Judith Weinstein, Theodore Goldfarb and Carol Cina

detailed accounts of a number of large-scale research projects that were underway then at Chungshan University and the Institute of Entomology in Canton, the Institute of Zoology in Peking, and the Shanghai Academy of Agricultural Sciences, including some that were initiated during that period.

One of the institutes that Needham singles out as having suffered most from the G₄ was the Natural Tropical Products Research Institute on Hainan Island. When we visited the Institute we found that most of the research was stopped in 1965 for eight years, that is, before the Cultural Revolution had even begun. Students and staff returned to the institute in 1973, and vigorous new research programmes started in 1975 (during the peak of the G's influence).

This example serves to illustrate a general observation we made during our visit. The Cultural Revolution clearly had a profound influence on the organisation and practice of science in China, and in many cases the level of scientific activity at research institutes and universities was slowed up or interrupted. Most researchers spent some time living in the countryside with the peasants and they devoted a considerable amount of time at their research institutes to political discussion. These policies were not instituted by the G₄ in order to punish scientists. Rather, they were principled

policy decisions made by the Chinese leadership to discourage the development of a scientific elite divorced from the needs of the people and to reverse the progress already made toward that end. (This does not mean that at the local level these policies were not occasionally used as a method of punishing particular individuals; we were told that this happened.)

The redirection of science and technology was initiated at the very onset of the Cultural Revolution in 1966. In certain respects, such as the demystification of science, the decentralisation of scientific decision-making and emphasis on the practical rather than the theoretical, the new policies were a continuation and intensification of those introduced in 1958 during the "Great Leap Forward".

What is more important than assigning blame (or acclaim) to individuals for a certain science policy is to analyse what the effects really have been and will be in the future. We learned that there is still a significant difference of opinion on the costs and benefits of the Cultural Revolution.

Many scientists we talked to thought that their experience of working in the countryside had been very valuable in developing a consciousness about the needs of the peasants, whom they are supposed to be serving. They developed ideas for their investigations by discussing their work with peasants, with whom they established continuing and mutually productive intellectual exchanges. For example, Li Li-ying, head of the Institute of Entomology in the outskirts of Canton, told us how the Laboratory of Biological Control gets most of its research ideas from nearby communes. Research begun there in 1974 is centred on ways to increase the natural population density of a predator which feeds on red mites, a common citrus pest. Tang Yuan, a weed control specialist at the Shanghai Academy of Agricultural Sciences told us that before the Cultural Revolution his research work, based on techniques developed in foreign countries, was rejected by the local peasants as unsuited to the conditions in Chinese paddy rice fields. When he began to work alongside the peasants and respect their knowledge about the problems that need to be solved he discovered a suitable herbicide which is now widely used in the Shanghai area. He now believes that intellectuals should integrate themselves with workers and peasants.

We also heard complaints however, from some scientists and technicians concerning their experiences during the



Manual work for students and scientists was policy not punishment

Cultural Revolution. They complained of serious interruptions to their research work by long periods of manual labour. In some cases no use was made of the training and skills of these researchers. We could not find out why the treatment of these individuals was so different from others.

Two universal complaints we heard were about the deemphasis of theory and a nearly total disregard for foreign scientific developments. While we are pleased with the reversal of these policies by the present Chinese leadership, our discussions raised some serious questions about potential problems that may result from the present drive to develop rapidly a large scientific and technical workforce as the key to modernisation by the year 2000.

First, we were told that a system of "key" schools (including secondary schools, universities, and agro-technical colleges) has been re-instituted. Set up at the national, provincial, and county levels, these schools are endowed with better facilities and faculty than most schools and their function is to channel the "brightest" students into the best schools in the hopes of rapidly producing a large work force of better scientists and technicians.

In addition, a system of nationally standardised exams was reinstituted this year. We were told that the results of these exams will be the most important criteria in deciding which students go to universities and which are sent to key schools. Further, the old requirement of students spending one or two years in manual labour before pursuing advanced studies has been dropped. Vice chairman Teng Hsiao-ping said in his 18 March speech, "Those who labour, whether by hand or brain, are all working people in a socialist society". Some scientists we spoke to

agree with this idea. Others, however, suggested that all intellectuals benefit from manual labour and contact with the masses. We also learned that the government has decided to give special perquisites such as higher wages and better living conditions to scientists.

Associated with these policy changes is the danger of the development of a scientific elite that will become separated from the masses and do work unconnected to the needs of the peasants and factory workers. During the years of the Cultural Revolution there was much concern about the possibility of such an elite developing in China, as had happened in the Soviet Union. The fact that China was a society in transition to socialism meant that social classes still existed. Before the Cultural Revolution most scientists and technicians came from the upper classes. Thus, at that time, the leadership saw the division of mental and manual labour as a form of class antagonism with the upper classes performing primarily mental labour and the peasants and workers primarily manual labour. Consequently, they emphasised the need for scientists and intellectuals to engage in political study and perform manual labour alongside the masses so as to better understand their living conditions and be reminded of who their knowledge and science should be benefiting. Virtually everywhere we went this summer we were told how important it was that scientists develop a proletarian outlook so that their work would serve the needs of the people, but some of the new policies seem to conflict with this goal.

Throughout our trip, however, we did see evidence of integrating science and technology with agricultural and factory work. An excellent example of this is China's Four-Level Agro-science Research Network. This network involves agricultural workers in research

and decision-making in agriculture-related science work. It consists of a county level research institute and a team of agricultural technicians and trained peasants for each county, commune, brigade, and production team⁵. The teams of technicians are engaged in agricultural research, agro-science, education, and popularisation of both the results and the methodology of scientific farming. We studied the network in detail in Wu county, Kiangsu province. Eighty per cent of the counties in China have been similarly organised since the networks were initiated in 1974.

We would like to emphasise that we did see considerable evidence that science will continue to serve the people of China. Yet, the political turmoil of the Cultural Revolution has subsided over the last several years. In its place, there has been an increased emphasis on stability and economic growth. Both Dr Needham and the editor of *Nature* apparently view this as a positive development: "So what do the Chinese scientists need most at present? They certainly need a period of many years of tranquility, away from the limelight of political attention . . . Science cannot survive intact given such a bumpy road". We are not so sure.

If one measures progress in science only in numbers of publications, research projects and highly trained scientists, perhaps they are right. At one time at least, however, the Chinese decided that this was not to be the measuring stick for their science. Rather they said that the best science comes from researchers who are fully integrated with other workers in the society and who are committed to serving the needs of the broad masses of the population. At times this may slow down the rate of scientific growth, as measured by Western standards. During the difficult transition to a socialist society, it may be that the "bumpy road" is the only way to achieve this alternative vision of science. Developments in China during the next few decades may help answer this question. □

Notes

1 Please address comments and questions to Science for the People, China Delegation, 897 Main Street, Cambridge, MA 02139. (The views expressed in this article do not necessarily represent the views of all members of the delegation.)

2 "The Chinese Economy in 1976", pp 362-4, 382, *China Quarterly*, June, 1977.

3 USCIA "People's Republic of China: Handbook of Economic Indicators", p 1, August 1976.

4 *China Quarterly*, note 9, June 1977.

5 A typical county has approximately 1 million people; a commune, 30,000 people; a brigade, 1,500 people; a production team, 200 people.

Frelimo concentrates on the practical side of science

Pamela Smith and David Wield recently spent two years in Maputo at the Ministry of Health and the University Eduardo Mondlane. Here they assess the state of science in Mozambique

MOZAMBIQUE, independent since June 1975, almost ten years later than most of her northern neighbours, spent this last decade of Portuguese colonialism fighting an armed liberation struggle. The necessity to gain independence through fighting Portugal, herself a backward country, had some positive aspects. By independence, for example, Frelimo policies were already clearly formulated and tried out in the 'liberated zones' of the north, which covered about one-third of the country's total surface area. These policies had obvious implications for higher education and research and as such for science.

Higher education was to exist therefore, not to train a small elite in advanced techniques, but to "permit access to scientific and technical knowledge by the peasants and workers, for them to be able to assume economic power and increase production for their own benefit." This statement was issued as part of a policy document emanating from a National Seminar on Technical Education, held in December 1975. It is further reinforced by the report of the Central Committee of Frelimo to the third congress in February 1977, stating that "In education, it is indispensable that we continue to promote a constant increase in technical and scientific knowledge among the labouring classes and guarantee their access to the higher levels of education." This position originates from early on in the history of Frelimo, when in 1968, the only Frelimo secondary school, based in Tanzania, was closed after students refused to go into the 'liberated zones' for their holiday activities, demanding instead higher education opportunities in foreign countries, far away from the realities of the armed struggle.

Policy in independent Mozambique continues to be formulated by Frelimo. It is then analysed at the planning ministry responsible, as its name suggests, for coordination and overall planning, and by commissions, including members of the various ministries, such as education, health, agriculture, industry, concerned with particular areas of study. Before any plan is made, discussion takes place at all levels, including the individual institutions, such as factories or schools where representative committees will have been elected. In a factory, for example, the workers elect a workers' council, which has res-

pensibility for suggesting measures to improve safety at work, or to increase productivity and quality of products. They may also make suggestions for the type of training courses required for different grades of personnel from inservice to undergraduate programmes. Proposals are then made on the basis of discussion reports received by the Planning Ministry and analysed once more at ministry level. Any criticisms or further suggestions are then directed back to the planning ministry. This two way process between the centre and the periphery is very important to keep policies based on everyday realities.

Mozambique is now coming to the end of the second year of her three year plan, outlined during the third party congress in February 1977. The plan's overall objective is to raise the economy to pre-independence 1973-levels, and to develop by 1980 the basic structure for the next period of national reconstruction. The plan, as it relates to science, is taking place on two levels: first in education, where as many people as possible are to be provided with the technical and scientific skills necessary to the development of the country; and second through research mainly in certain priority areas, such as geology and crop production. Unlike many other third world countries, Mozambique does not possess a national science policy research council. Science policy is seen as an integral part of an overall plan, and as such, statements related to research can be found in the economic and social directives in the third party congress documents.

How can the two level objective be put into practice, given the colonial education system which gave only 5-10% of Mozambique's population the skills of reading and writing? Or which set up a few luxury secondary schools and a university in Maputo (formerly Lourenço Marques) for the Portuguese settlers and a handful of African and mixed race 'Assimilados'? Research done at the university and government research institutes, was for the most part irrelevant to Mozambique's needs and only questionably useful to the colonial power herself. Portugal, though poor and backward as a nation, under the Salazar and Caetano regimes, was bent on the pursuit of a high level of research to gain international pres-

Sorry, for copyright reasons some images on this page may not be available online

Camera Press

tige. Even in the Mozambican agricultural sector, with large exports, research was very restricted in comparison with neighbouring colonies like Tanganyika, which at least had competent research programmes and well staffed institutes for the major export crops such as sisal, coffee and cotton.

The president of Mozambique, Samora Machel, said when visiting the university in May 1976, "It is not by chance that we find that our university engineering graduates working in generating stations cannot begin to repair a generator, and medical professors who have difficulties diagnosing simple illnesses of the local population". Not only was the top scientific education that Mozambique had to offer at independence virtually useless for national reconstruction, but the schools' system was also highly theoretical and restricted to a minority, leaving an illiteracy rate of over 90%. Only those in the elite 'liceu' secondary schools could enter the University. Those in the industrial and commercial secondary schools, had to be satisfied with technician training in lower level technical institutes.

It is for this reason that in the early years of independence priority has been given to building the educational base, particularly in the rural areas. Crash literacy programmes have begun using not only school and university students but also those who are themselves barely literate, to teach those who are not. The primary education system, which takes four years to complete, has trebled since 1974 and now about 1.2 million pupils are registered.

New schools are often set up through local initiative, and in the absence of buildings pupils can be seen learning beneath the trees of many Mozambican villages.

At independence, the secondary schools shrank rapidly as the Portuguese settlers left, but there are now over twice as many schools and pupils as in 1974. In the university, founded in 1962, the student population of over 2,000 in 1973 diminished to one-quarter of this number in the early independence period. University teachers and technicians also left in great numbers, being part of the settler and 'assimilado' population, with the result that faculties and research institutes were denuded of qualified personnel.

Further data on scientific training, manpower and budgeting is difficult to obtain. Colonial statistics were either incorrect—for example, the last population census in 1970 underestimated the population by over 20%—or non-existent. Mozambique is one of the few countries in the world with no government estimate of gross national product. However a national statistical centre is in formation this year, and a national accounting system will be organised by 1980. More comprehensive educational statistics are being collected this year but are not yet available.

Nevertheless it is possible to estimate that there are now 200-400 Mozambican graduate-level scientific workers, mostly involved in agriculture, health, education and engineering. The majority of these are under 30 years old, and were trained in the pre-independence period. They therefore had to make a positive decision to remain in Mozambique, helping to keep services running amid the initial confusion that ensued as many of their colleagues left for Portugal. There are a few Mozambicans with postgraduate qualification from Portugal, South Africa, England, France and USA. As the policy of the government was to keep the university and research institutes going, these Mozambican scientists find themselves working alongside a considerable number of foreign scientists, some of whom remained from the pre-independence period to help national reconstruction, as well as those recruited since 1975 from socialist and progressive countries and groups.

The university and research institutes work under the general orientation of the ministry of education, placing their 'human and material resources . . . at the service of the social and economic priority sectors', as defined during the third congress. A scientific worker may therefore find him/herself engaged in curriculum



Literacy is the educational priority, particularly in rural areas

planning, teaching university students and technicians, organising fieldwork and research projects and serving on interdisciplinary commissions at ministry level. Within the university, some changes have been effected. Courses have been reduced from five or six years' duration to three. In 1975, for example, the final year agronomy students decided that they did not wish to graduate until they knew something of the agricultural reality of the country. Instead of graduating, they volunteered to work in the countryside for a period. Other courses have since become more practically orientated. Last year, for the first time, chemical engineering students learned control engineering and smelting practice, in local factories as well as in the classroom. Previously, they would have conducted routine experiments solely within the laboratory and studied theoretical mathematics and physics. Hence for example, complex atomic theory has now been withdrawn from the curriculum, permitting a greater emphasis on more relevant theoretical and practical work.

In addition to existing courses, university teachers have been asked to set up new ones for lower level students. Some of the teachers were unhappy, but the majority perceived a way of changing elitist attitudes in the university. Pre-university courses for workers and crash teacher training programmes have begun. Only about 20% of the students are fulltime, the others provide technical backup in the ministries, schools, factories and other government institutions. The university's function was described by the president as having "to leave its doors and go to the factories and villages, putting its techniques at the service of national reconstruction. There it will learn new techniques from practical life. This will stimulate the study of scientific realities, not in an abstract way, but in a creative manner, linking its knowledge to the material wellbeing of the population."

New research priorities and methods

of working have thus emerged. As already indicated, there is a close link between the university and the various ministries, particularly in the field of scientific research. Scientists are working in the ministry of health and the faculty of medicine; others are based in the agronomical and veterinary faculties, the ministry of agriculture and two research institutes in Maputo or small up-country laboratories.

One specific example of the type of research project that is being undertaken in Mozambique, is in the field of traditional medicine. Two hundred plants have already been collected which traditional healers claim have medicinal properties. Library research is currently being undertaken by a commission in the ministry of health, to see what is known about them. It is already clear that most traditional treatments for diarrhoea contain tannin. Within a short period, some plants which the literature search suggests are safe, will be tried out, not by extracting the active ingredients and turning them into modern pharmaceuticals, but by encouraging rural health workers to use them in the traditional manner, such as in tea form.

In the university, research has been geared to seeing how technology can be adapted and used according to priority needs. One final year student of chemical engineering carried out an important project on charcoal production, charcoal being the fuel used by the majority of Maputo's 800,000 population. Deforestation of the surrounding areas of the city seems to be one of the problems resulting from present production methods. Her work was to look at these methods not only to see how to improve their efficiency, but also to check on possible extraction and use of methanol and acetic acid from wood distillation, since Mozambique does not produce either of these chemicals. She was also investigating the effect of centralising production on family level charcoal producers, dependent on this for their livelihood. This

project has since been integrated into the planning of a larger project, based on the ministry of agriculture's forestry department. Also within the university, an appropriate technology unit has been set up, working on problems such as methods of grain storage, food preservation and water purification, especially relevant to the rural areas. In the ministry of agriculture, soil studies have begun as have studies of local foodstuffs and their usefulness for poultry rearing.

Sometimes, useful research results come unexpectedly. A recent study on the extent of migrant labour from Mozambique showed that returning labourers have technical skills which are in short supply. Another example is the national vaccination campaign, which is almost complete, having begun in 1976. Mobilisation has been extremely good and WHO estimate that already 92-96% of the population has been vaccinated against smallpox, tuberculosis, DPT (Diphtheria, pertussis, tetanus) or measles, depending on age. The 1970 census figure of 8.3 million has already been exceeded and estimates expect final totals to be in the region of 11 million. The vaccination campaign also gives demographic information on the age distribution and density of the population. In the past four years, Mozambique has been able to restructure its educational system placing the main emphasis on basic skills. Its higher scientific manpower, small in number, has been actively involved in this restructuring in teaching, planning as well as research. Teaching has become more practical, planning has had to take into account national priorities and research has thus been research with a small 'r', requiring only restricted laboratory and library facilities and using students as well as teachers.

This year, one of the priorities of the university and government research institutes has been to plan for research in the medium term that is for the next planning period beginning 1980, not as until now, on a short term basis. Stricter planning is required and the lessons learnt from the last few years, are being analysed in detail so that improvements can be made ensuring that research tasks are put more, not less, into line with national priorities.

The Mozambican government continues to recruit scientists able to work in various fields. Those who feel that they may have relevant skills and could be useful in the Mozambican context, should write to: — The Mozambique Information Centre, 34, Percy St., London W1P 9FG, Telephone No. 01-636 7108 who coordinate recruitment in Britain and would be happy to receive enquiries. □

'By nature not nurture' they sang . . .

AT THE recent International Conference on the Unity of the Sciences in Boston in the United States of America, the life sciences committee, of which I was chairman, had a useful, calm and objective discussion on genetics and society, when the importance of inheritance and environment in man was considered. One of those present was Professor Arthur Jensen from the Institute of Human Learning of the University of California. At the end of the meeting he said that he had been very favourably impressed by the behaviour of the discussants and of the audience; all shades of opinion had been expressed without fear of interruption. He compared this experience to that obtained before audiences in universities in America and Europe, where certain left-wing groups have disrupted the meetings, saying that these topics should not be investigated or publicly debated.

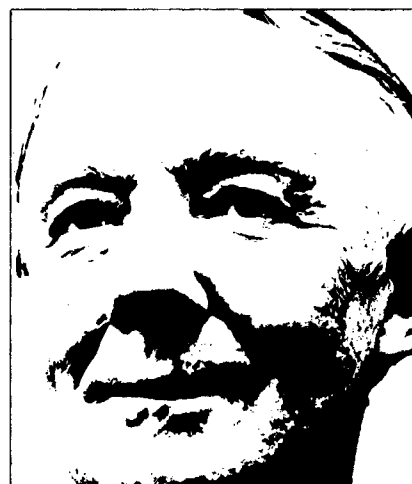
It is interesting to note the changes which have occurred in the attitudes of the politically motivated in this field. In the 1920s there were many Socialist Sunday Schools in Glasgow, Scotland, where young people were nourished in what was then considered to be pure Marxist doctrine. Techniques of instruction familiar to the audience, many of whom had previously attended Church Sunday schools, were used. They listened to "sermons" and sang secular "hymns". One favourite hymn included the lines:

"By nature not nurture we'll rise to the skies

The means of production we'll nationalise."

Although nationalisation may still remain within socialist political programmes, views on nature and nurture appear to have changed. Fifty years ago those running the Sunday schools took the view that those of their pupils with innate, inherited ability would triumph over the effects of their unfavourable environment. In fact they believed that the effects of heredity could be even more important than those of the environment, if the individual then received the appropriate type of education. Professor Jensen and others have, in recent years, been vilified for suggesting that this may still apply.

There has also been a change in attitude to the use of Intelligence Quotients (IQs). Left wing opinion



KENNETH MELLANBY

condemns these as instruments in promoting racism. However, at one time these were considered the fairest method by which bright children from poor educational backgrounds, who found conventional examinations difficult, might be admitted to grammar schools. The intention was not to show that the underprivileged were stupid, but to give them a better chance to develop their potentialities.

I found IQ tests to have some value when selecting students to admit to the first university courses we started at Ibadan in Nigeria in 1948. Our infant university could only admit a very small proportion of those who applied. Nigerian schools at that time were very variable, ranging from the government colleges which compared with grammar schools in Britain, to those run privately with few trained staff. We set what were then considered conventional written papers in arts and science subjects, and I also gave each candidate an IQ test. Most entrants were selected on the results of the written papers, and none whose marks were above an arbitrary minimum was excluded, but I also admitted a few who did badly in written work but who had high IQ scores.

I am glad to say that members of this latter group almost all did well in their later university courses. In fact I found that, for all students, their results in their degree examinations correlated better with their IQ figures than with their marks in the entrance examination. Incidentally the IQ scores of our successful Nigerian students were in the same range as those found with white undergraduates I tested in Britain, so these findings should not be offensive to those who, for political reasons, oppose this type of investigation. □

correspondence

Biochemical nomenclature

SIR,—Every biochemist knows the frustration of being told to use a new name for a substance when he was perfectly happy with an old one. It can, however, be just as frustrating when different groups of workers use different nomenclature systems. Widely agreed recommendations may, at the very least, prevent a name from implying an incorrect structure; they may even indicate correct ones.

Slavish adherence to fully systematic nomenclature gives unusably long names; at the other extreme laboratory jargon is understood only by the initiated. There is therefore a need for specialised nomenclatures in many fields to compromise between the needs of the workers in the field and more general readers.

A difficulty in the past has been that names may be established before it is realised that they are misleading or ambiguous to those in related fields. A glaring example of this is 'diphosphate'. This was long used to mean two phosphate groups, and also, as in ADP, a single diphosphate group. The latter usage has been approved, and two separate groups are now distinguished by being called 'biphosphate' [strictly bis(phosphate)] as in fructose 1,6-biphosphate.

Much of the difficulty can be avoided if recommendations for a self-consistent nomenclature are agreed at the very beginning of the study of a new class of compounds. This is the thought behind general principles for naming natural organic compounds, prepared by the Commission on Nomenclature of Organic Chemistry of the International Union of Pure and Applied Chemistry (IUPAC) and published as: *Nomenclature of Organic Chemistry, Section F, IUPAC Information Bulletin, Appendix 53, 1976 (Eur. J. Biochem. 86, 1–8; 1978).*

The basic idea of Section F is to have a set of stem names for each group of natural products to which the general principles of organic nomenclature will apply. Thus, unsaturation and the presence, type and number of functional groups will be indicated by prefixes and suffixes according to the standard methods of Nomenclature of Organic Chemistry, Section A–D, stereochemistry according to Section E and isotopic substitution according to Section H. Modifications of a skeleton will be shown by methods first introduced in the Steroid Rules, and now set out generally in Section F.

Although Section F is designed to guide individuals so that they can themselves propose a clear and simple system of nomenclature for a new class of compounds, there remain great advantage in bringing together many workers in a field to agree on how its general principles should be applied. The Joint Commission on Biochemical Nomenclature (JCBN), joint between IUPAC and IUB (International Union of Biochemistry), set up in 1977, has, as one of its main tasks, the application of these principles to appropriate groups of natural products. In this it works closely with the

Nomenclature Committee of IUB (NC-IUB).

In some areas recommendations for nomenclature have already been made. In others JCBN will be inviting specialists who are active in the biochemistry and chemistry of particular groups of compounds to form working parties to make such recommendations.

We hope that our colleagues will accept invitations to serve on such working parties and carry out an important task. We hope too that specialists in areas that need agreed nomenclature will offer to form working groups themselves. We would be glad to hear of suggestions for topics from any groups in such areas that already meet for discussion. We would be glad to hear of suggestions for topics that may need nomenclature agreements, even when not accompanied by an offer to serve on working parties!

P. KARLSON

*Chairman, JCBN & NC-IUB,
Institut für Physiologische Chemie,
Marburg, West Germany*

H. B. F. DIXON

*Secretary, JCBN & NC-IUB,
Department of Biochemistry,
University of Cambridge, UK*

We gratefully acknowledge the initiative of the late W. Klyne in preparing this letter.

Has food production kept up?

SIR,—John Gribbin has referred (16 November, page 208) to FAO data indicating that in the past two decades, while population has grown at 2% per annum food production has grown by an average of 2.8% per annum. As I have pointed out elsewhere (*Proc. Nutrition Soc.* 36, 121; 1977) such aggregate figures are highly misleading. Whereas, in the developing countries, the total increase in food production was 2.8% per annum in the 1960s, demand for food increased by 3.5% in those countries. Almost two-thirds of all developing countries showed a fall in self-sufficiency in 1970–72 as compared with 1961–63.

Even more critical is the situation in the food-deficit less developed countries, in which grain production increased by 2.5% per annum in 1960–74 but only by 1.7% per annum in 1967–74. Analysis by the International Food Policy Research Institute shows that even if grain production in those countries were to increase by 2.5% per annum to 1985, they would by then have a deficit of 100 million tonnes of grain per annum. The trend over the past four decades has been of an increasing surplus of grain in North America and an increasing deficit in the developing countries. The problem, therefore, is not that of overall world food production but of where the food is produced, and of its distribution, both between and within different countries, and also whether it is used to feed animals or human beings.

TILO L. V. ULBRICHT

*Agricultural Research Council
London*

Mr Moon's philosophy

SIR,—While I know very little about the ICUS of 1972 and 1973 and the organisation of Mr Haskell, it is ridiculous for Professor Nicholas Kurti (16 November, page 206) to place the ideas of Mr Moon on a par with those in Haskell's book "Full Circle: The Moral Force of Unified Science" where the latter postulates periodical tables for botany, zoology and sociology.

Professor Kurti seems to assume that he can simplify to the necessary degree to reach his conclusions situations which are extremely complex. The 'Give and Take' relation is a fundamental metaphysical principle and we are consciously or unconsciously dependent upon it. Mr Moon's philosophy asserts that the true unity exists in the Creator while in the creation there is only plurality even though it is ordered plurality. Mr Moon is concerned primarily with the order of life and with ordering life and he presents us with a conceptual apparatus which sets out many things to which we can relate in a gradual and ordered way.

I must explain that the ICUS meetings are not public relations ploys to further the Unification Church, nor are they frosting on Mr Moon's theological cake. Mr Moon founded the ICUS conferences because of his deep and sincere desire to set up a forum to further the striving of man to resolve "the most important problem of our time—the reconciliation of religion and science. The problem is not simply a matter of reconciling two academic disciplines. Rather the problem points to the need to investigate the dichotomy of facts (science) and values (religion and ethics). This separation of science and religion in the modern world has caused untold harm and serious damage to the development of humanity."

But while Professor Kurti himself says that the ICUS conferences should be judged by the contents of the published conference proceedings and not by selected remarks and others, his ideas about the teachings of Mr Moon's movement indeed echo the popular press. It would have been better for all concerned if Professor Kurti had consulted the official publications of Mr Moon's teachings and the activities of his movement.

JANE SALTER

*International Cultural Foundation,
44 Lancaster Gate, London W2*

Laser progress?

SIR,—The Christmas laser display in Oxford Street, London has just given rise to the following exchange of conversation:

Scientist: "I have a laser almost as good as that in my lab."

Companion: "Yes! But is it in colour, or only black and white?"

C. B. LUCAS

University of London, UK.

news and views

Echoes of the bat's cry

from Jennifer Altman and Shin-Ho Chung

As the basic mechanisms of hearing are beginning to be understood, auditory physiologists are turning their attention to the neural processing of complex acoustic signals. From several recent studies on amphibian (Pettigrew *et al.* *Nature* **272**, 138; 1978), avian (Knudsen & Konishi *Science* **200**, 795; 1978) and mammalian (Chalupa & Rhoades *J. Physiol. Lond.* **270**, 595; 1977) species, we now know that auditory space is accurately mapped onto a nucleus in the midbrain, while neurones in the higher cortical centres are specifically tuned to extract certain features of sound. As might be expected, those animals which rely mainly on their ears for survival have the most highly developed auditory brain, and slowly neurobiologists are unravelling the neural mechanisms by which such animals handle acoustic signals.

Of the several known species which construct a picture of the external world by listening to echoes from ultrasonic pulses, bats are the best known. Each species of bat has its characteristic ultrasonic cry. Some, such as the little brown bat (*Myotis lucifugus*), use cries which are frequency-modulated, changing in pitch from high to low, whereas others, such as the moustache bat (*Pteronotus parnellii rubiginosus*), emit a cry of constant frequency, which tails off with a brief downward change in pitch. As the animal approaches the target, the emission rate of its cries increases, and the duration of each cry is shortened. From the echoes of its cries, the animal then gauges the size, distance and velocity, as well as the location, of the object ahead of it. Such an unorthodox way of seeing the external world is extremely efficient, as demonstrated by the agility with which bats avoid moving barriers (Jen & McCarty *Nature* **275**, 743; 1978),

and the accuracy with which they catch their prey. It is estimated that a bat on a hunting spree catches about 500 insects per hour.

Several recent studies on the horse-shoe bat (Möller *et al.* *J. comp. Physiol.* **125**, 217, 227; 1978) and on the big brown bat (Feng *et al.* *Science* **202**, 645; 1978) show how cells in the mid-brain nuclei respond immediately to the distance of an object gauged by echolocation. In the auditory midbrain, there is a class of neurones which discharges only when an ultrasonic pulse mimicking the animal's cry is followed by a faint echo. These neurones are particularly sensitive to the delay between the pulse and echo, and different neurones are tuned to different ranges of pulse-echo delays; the optimal delay for eliciting discharges varies from 4.5 to 20 ms, which corresponds to target ranges from 0.75 to 20 m. Indeed, the entire midbrain appears to be elaborately designed to extract information from a pair of stimuli. Möller (*op. cit.*), for example, showed that the presentation of a first tone briefly enhances the sensitivity of the midbrain neurones, so that a faint echo following soon after can effectively be detected. The precise excitatory and inhibitory neural network underlying these observations remains to be elucidated.

In a series of studies on the moustache bat, Suga and his colleagues have demonstrated that a large part of the auditory cortex is devoted to detecting a slight shift of frequency components of the returning echo compared with the emitted cry (Manabe *et al.* *Science* **200**, 339; 1978; Suga *et al.* *Science* **200**, 778; 1978; Suga *Fed. Proc.* **37**, 2342; 1978). This Doppler shift contains information about the relative velocity of the animal and the target, and a remarkable behavioural adaptation enables the bat to utilise the specific tuning of its auditory neurones to the full. When the frequency of the returning echo is shifted to, for

example, 63 kHz from the emitted frequency of 61 kHz, the bat reduces the pitch of its cry, so that the Doppler-shifted echo frequency would be at 61 kHz, the frequency to which the inner ear is specifically tuned (Suga & Jen *J. exp. Biol.* **69**, 207; 1977) and to which a disproportionately large area of cortical surface is devoted. In the primary cortex, neurones responding to frequencies from 61 kHz to 63 kHz are ordered concentrically, with the area excited by a 61 kHz tone located at the centre and the surrounding rings of cells excited by progressively higher frequencies. In addition to frequency selectivity, each neurone in this area responds only to a very restricted range of sound intensity. Lines of isoamplitude are at right angles to the iso-frequency contours, and each neurone has both a best frequency and a best amplitude. When the echo is faint, neurones located on the lower part of the auditory cortex respond and these are excited by inputs from both ears. As the intensity of the echo becomes louder, neurones located on the upper part of the cortex begin to respond. These units are directionally sensitive, as they are driven by an input from the contralateral ear and inhibited by the ipsilateral ear. In this way, the brain encodes the location and the velocity of a target.

Further towards the centre, there is a cortical region which is devoted to processing frequency-modulated signals. Unlike the Doppler-shift region, this portion of the auditory brain has only a vague tonotopic representation, but contains instead clusters of neurones with similar response characteristics. The stimulus parameters to which neurones respond are more specific. Some neurones, for example, respond best when the frequency-modulated sound is followed by a faint overtone of the first sound. Others respond only to a pair of sounds with particular combinations of frequency, amplitude and timing, similar to the delay-specific

Jennifer Altman is a lecturer in Zoology, University of Manchester, and Shin-Ho Chung is at the National Institute for Medical Research.

cells in the midbrain. These neurones thus seem to act like feature detectors, and by comparing the spectral composition of echoes with the emitted signals, they may serve to detect the size and location of the reflecting object.

The study of the bat's auditory system is providing us with unexpected insights into how the brain is designed to utilise biosonar signals. The next task for researchers is to elucidate the intricate neural network that enables

certain features of acoustic stimuli to be extracted, and the processes by which such connections evolve during development. Moreover, it will be of great interest to know how the middle ear of the animal is constructed so as to vibrate optimally at ultrasonic frequencies. Once this task is achieved, we shall be a step nearer to understanding the mechanisms of sensory processing and integration, for which the brain is most exquisitely designed.

Positrons at surfaces

from John Pendry

RADIOACTIVE sources provide a supply of positrons in the laboratory, and one of the applications of these particles is to studies of the solid state. In a recent Letter (Mills, Platzman & Brown, *Phys. Rev. Lett.* **41**, 1076; 1978) an important new advance in technique has been demonstrated: positrons can be controlled precisely enough to probe the surface of a solid on an atomic scale.

Conventional experiments use positrons straight from the source with energies of the order of an MeV, and fire them into the solid under study. Once inside, a positron interacts strongly with the electrons through Coulomb forces. Plasmons, collective oscillations of the electrons, rapidly dissipate the energy and in a very short time the positron has lost almost all its energy and momentum. Capture of the positron by its antiparticle, the electron, will eventually take place, producing a pair of γ rays with the disappearance of the positron and an electron, though the time for annihilation is long compared with the time to come to rest in the solid. By conservation of momentum the γ rays must have almost oppositely directed momenta, slight deviations from 180° being due to the relatively small but finite momenta of the electrons with which the positrons annihilate. Measurement of the deviation, in experiments which look for coincidences of two γ rays, gives the momentum distribution of electrons in the solid, information which can be used to understand bonding of solids, and to measure the Fermi surfaces of metals.

An additional complication arises for insulators, where the positron's Coulomb potential is largely unscreened, and is strong enough to bind an electron into a hydrogenic state known as positronium. The positron will preferentially annihilate with its

'own' electron, and signals characteristic not of the solid, but of the positronium bound state, are seen.

Surface sensitivity in positron experiments came a step nearer when it was shown that the thermalised positrons can be re-emitted from a solid before they have time to annihilate. Unlike electrons, positrons tend to be repelled by many solids and if in diffusing around the solid they encounter a surface they will be expelled from the solid. The positrons do not acquire more than about 1 eV energy so are reasonably monochromatic and can be accelerated by electric fields to form a collimated beam with easily controlled energy (Canter, Mills & Berko *Phys. Rev. Lett.* **33**, 7; 1974).

The point is that positrons with only 100 eV kinetic energy have limited powers of penetration, typically a few tens of Å, and therefore can be used to probe surfaces on an atomic scale.

Mills, Platzman and Brown have performed the first experiments of this nature. They worked with clean, well characterised, single crystal surfaces of aluminium, silicon and chromium. The characterisation by LEED and Auger techniques is most important in surface experiments as contaminants can easily cover a surface and obliterate all of the clean surface properties. The energies of positrons re-emitted from the surface were analysed by applying a retarding field to the sample and measuring the increase in the number of positrons annihilating inside the sample. The spectra had two components: a sharp peak at very low energies due to positrons that had thermalised inside the solid, and a much broader distribution due to incompletely thermalised positrons. The latter component became more intense as the incident energy was lowered below 1000 eV, a consequence of decreasing penetrating power.

The sharp peak was seen in all the materials studied, beginning its rise at zero kinetic energy and having a width

dependent on the solid: 0.5 eV for aluminium, 1.0 eV for silicon and 1.7 eV for chromium. Because the positrons in this peak have thermalised inside the solid, the peak cuts off at a maximum energy which is the 'negative affinity' the solid has for positrons.

Our understanding of positrons in solids and at surfaces is based on calculation and on some experimental annihilation rates and angular distributions. The new experiments will provide a more detailed picture of positron behaviour. Theories of the positron-solid force law such as that of Hodges & Stott (*Solid State Commun.* **12**, 1153; 1973) can now be checked, even to the extent of diffracting positrons from surfaces. The rate of energy loss of positrons in solids can be determined from the ratio of thermalised to unthermalised positrons emitted: accurate calculations are difficult especially when the positron is moving slowly. Some new problems present themselves. The width of the low energy emission peak is only of the order of 1 eV but this is still more than an order of magnitude larger than the thermal spread in energies. Simple theoretical arguments require that the positrons are accelerated normal to a perfect surface, and if inelastic collisions are excluded, emerge in a parallel beam with only a thermal spread of energies. Certainly this is the situation found for electrons in negative electron affinity systems. The emission process for positrons must be more complicated and is yet to be elucidated.

Finally, we can look forward to the more difficult experiment of detecting emission of positronium. Canter, Berko and Mills (*Phys. Rev. Lett.* **33**, 7; 1974) have shown that a large fraction of the positrons pair with electrons as they leave the solid and emerge as positronium. This pair-tunnelling process reminiscent of the interface between a normal and superconducting metal can be studied by measuring the angle and energy distribution of the ejected positronium. □



A hundred years ago

THE *Japan Mail* states that an astronomical observatory is to be established within the precincts of the Geographical Bureau of Tokio. The same journal also announces that telegraphic insulators, made at the village of Imari, in the province of Hizen, are of such good quality that they find large sale in Europe.

From *Nature* **19**, 19 Dec., 159; 1878.

John Pendry is at the Daresbury Laboratory of the Science Research Council.

Discrepancies in claims for endothermy in therapsids and dinosaurs

from R. E. H. Reid

L. S. RUSSELL'S claim, made in 1965, that dinosaurs were warm-blooded animals, has since been supported chiefly by the work of de Ricqlès on bone histology, and in papers by Bakker on other topics. The concept of dinosaurian endothermy has been welcomed by some authors, and represented as established in a popular book by Desmond; but such workers as Feduccia, Thulborn and Charig have remained unconvinced, and have severely criticised some of Bakker's claims. As usual some disputed issues authors are mistaken, while some of identical evidence, but one of Bakker's later arguments involves a serious factual error, unless many other authors are mistaken, while some of the arguments used in different papers are inconsistent or contradictory.

Bakker's later views appear in summary in his popular article 'Dinosaur Renaissance' (*Scient. Amer.* 232, 4; 1975), in which late (Middle and Upper) Permian reptiles from southern Africa are said to be exclusively therapsids. These are interpreted as endotherms, and as living in cold post-glacial conditions, which restricted large ectotherms to the Permian tropics. They are restored as possessing hair as insulation, and as able to stay active in winter snows. Even Triassic climates were supposedly not warm throughout the year, and large thecodonts (erythrosuchids) from the early Triassic of South Africa were therefore probably also insulated endotherms. The relevance to dinosaurs is that thecodonts are the group from which dinosaurs originated.

These claims are consistent with de Ricqlès' conclusions from bone structure (in *Morphology and Biology of Reptiles* (eds Bellairs & Cox), 123 (*Linn. Soc. Symp. Ser.* 3, no. 3, London, 1976)), but, unfortunately, not with other evidence. Charig points out (*ibid.* page 65) that glaciation of southern Africa was not Permian but late Carboniferous, and that the late Permian reptiles include various non-therapsid types, some of which (pareiasaurs) reached large sizes. There was even a large labyrinthodont amphibian (*Rhinesuchus*). This information is easily available from well-known textbooks, such as Romer's standard *Vertebrate Paleontology* (third ed., Chicago University Press, 1966) or Haughton's *Geological History of Southern Africa* (Geol. Soc. S. Afr.,

1969). Even stranger, Bakker's earlier analysis of predator-prey ratios, in a paper published in *Nature* (238, 81; 1972), shows more than one-third of the total prey biomass for the earliest late Permian zone (*Tapinocephalus*) as consisting of pareiasaurs, and describes its therapsids (deinocephalians), with pareiasaurs, as "almost certainly ectothermic". Later treatment of early therapsids as endothermic is within the bounds of reinterpretation; but what has become of the pareiasaurs?

Bakker also gives two opposite pictures of gait in therapsids. In his study of tetrapod locomotion (*Evolution* 25, 636; 1971), he insists that their gait was always sprawling, and never even semi-erect; but, in 'Dinosaur Renaissance', even the earliest are said to have limb modifications which permitted trotting gaits. This leads to a further problem. In the earlier paper, Bakker argues that the 'fully erect gait' seen in dinosaurs is probably necessary for large endothermic tetrapods, because of the food-gathering requirements of endothermic physiology, and claims that the echidna *Zaglossus*, at 10 kg weight, "illustrates about the size limit possible for a homeothermic vertebrate with sprawling limbs." Large therapsids which he estimates at weights between 500-1500 kg in 'Dinosaur Renaissance' should hence have needed erect gaits as endotherms, if this argument for dinosaurian endothermy is justified, but they cannot have had erect gaits if his analysis of limb movements is correct. In any case early forms seem unlikely to have had them.

The type of bone called 'endothermic' by Bakker is compact bone with a primary structure called fibrolamellar by de Ricqlès (often, laminar bone *sensu* Currey), with more or less extensive Haversian substitution. This is now characteristic of large endotherms, and is seen by de Ricqlès (*op. cit.*) as reflecting their rapid initial growth rates, which in turn reflect high metabolic rates. A possible alternative correlation would be with size and gait only. It might, for instance, be claimed that bone formed from densely packed cylindrical primary osteons, with progressive Haversian remodelling as size and weight increase, is simply better adapted mechanically than lamellar-zonal ('ectothermic') bone as a structural material for the limb bones of large animals whose limbs support weight for long periods. It may even be needed for them to reach large sizes. This seems ruled out, however, as a primary correlation, if therapsid gaits

are pictured conventionally, by the occurrence of this type of bone in early deinocephalians, and even the eotheriodont *Biarmosuchus*, but, ironically, this does not follow if even early forms possessed erect trotting gaits.

Treatment of therapsids as endotherms also weighs against this status for dinosaurs. Feduccia argued that impetus towards endothermy will have been greatest during periods of severe climate, and insufficient to lead to it in the times in which dinosaurs originated. Dodson's reply that "the ancestors of the dinosaurs lived at the same time and in the same place as the predecessors of the mammals" is valid if mammalian endothermy first arose in the immediate ancestors of the earliest 'true' (osteological) mammals. However, the precursors of the earliest therapsids were Lower Permian spenacodonts. The earliest known thecodonts only date from the end of the Permian, long after large pareiasaurs had been able to enter southern Africa; dinosaurs did not appear until the Upper Triassic, when conditions presumably approached those of the Lower Jurassic optimum. Bakker claims Mesozoic climate as optimal at the end of the Jurassic, but marine invertebrates show minimal latitudinal zonation in Lias times, and a 'boreal' realm persisting from Bajocian times onwards.

It is not hard to find other minor inconsistencies. For instance, an estimated maximum sustained speed of 2.8 km h⁻¹ for ostrich dinosaurs as ectotherms is contrasted with their having limbs like those of ostriches, which reach 50-80 km h⁻¹ (*Nature*, 238, 81; 1972). Limb proportions however, are an index of speed, not endurance. In 'Dinosaur Renaissance', even early therapsids are supposed to have trotted, but examples (pristirognathids) are shown with semi-erect limbs, which would not permit trotting. Dinosaurs found within the Cretaceous Arctic circle in Canada are seen as suggesting the ability to tolerate cold, but large Cretaceous crocodiles from Saskatchewan are later pointed out as ectotherms. Such dinosaurs would also have needed to be migratory, unless able to feed in winter darkness. The predator-prey ratio for the early Cretaceous Cloverly Formation is based on the assumption that a 70 kg deinonychiid preyed on two dinosaurs over 28 times heavier (2,000 kg) than itself, one of which was heavily armoured, and on a brachiosaur 571 times heavier (40,000 kg). This seems quite unlikely, unless it was simply a

R. E. H. Reid is in the Department of Geology, The Queen's University of Belfast.

scavenger, whereas, the specialised claws of *Deinonychus* suggest active predation on animals about its own size.

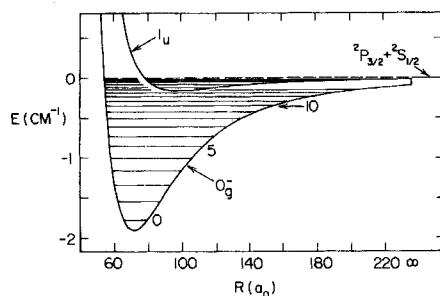
How do such problems affect the concept of dinosaur endothermy? They do not affect other arguments, such as Ostrom's contention that erect gaits may only be possible for endotherms, or Wheeler's recent comments (*Nature* 275, 441; 1978) on cooling of the central nervous system. Erythrosuchids need not have been insulated, but did have whatever condition is responsible for 'endothermic' bone. The problem is that doubtful claims tend to discredit real evidence. But even so, if one follows Charig's rejection of predator-prey ratios, the only major line of evidence remaining is that from bone histology, which cannot be certainly attributed to endothermy rather than bulk.

I must also support Thulborn's view that small size, a naked skin, and endothermic physiology seems a lethal combination for hatchlings. This need not mean that dinosaurs were simply 'good reptiles', but their nearest possible approach to true endothermy, if their young lacked insulation, will have been some combination of fairly high metabolic rates with poikilothermy, (until temperature could be stabilised by bulk in the larger animals). Perhaps this type of physiology would be feasible for typically large animals, evolved in warm Middle to Upper Triassic conditions, but it does not correspond to the fully developed endothermy of mammals and birds, in which external insulation allows high stable temperatures to be maintained in small animals. □

Long-range molecules

from Graham Richards

VAN der Waals forces are normally thought of as attractive. They are the weak attractions which hold molecules together with a binding energy often following the sixth power of the intermolecular separation. In fact this is only true for spherical charge distributions. Attractions of the c/r^6 form are appropriate for interactions between rare gas atoms but in more complicated instances this may not be true even between atoms. Perhaps most striking is the fact that if one of the atoms is in a spherically symmetrical S state and the other in an electronic state of P type then not only is the dispersion force of the form c/r^3 but



The potential-energy curves for two long range states of Na_2 .

it is repulsive!

This fact is utilised in a description of a new type of molecular electronic state by Stwalley and Yea-Hwang Uang of the University of Iowa and Pichler of the Joint Institute for Laboratory Astrophysics in Boulder (*Phys. Rev. Lett.*, 141, 1164; 1978).

The energy of a diatomic molecule is normally represented by a potential energy curve with a left-hand limit showing the vibrational turning point with the nuclei close together, and the right-hand branch the maximum stretch between the nuclei at the other extreme of vibration. Vibrational levels near dissociation give rise to 'long-range molecules' since although the molecule repeatedly vibrates classically from the inner turning point (typically 1 Å separation) to the outer one (typically 10–100 Å), it spends most of its time at the long range region.

The new type of pure long-range molecule is predicted to spend its entire time in long range regions with separations between 30 Å and 100 Å. These long-range molecules will be possible for electronic states where the repulsive van der Waals forces at large internuclear separations result in there being a small bump in the potential energy curve and a very small well in which vibrations can take place. The local energy minima at very long range arise from the recoupling of angular momenta.

In the familiar type of diatomic molecule potential curve the range of internuclear separation shown is perhaps 1–10 Å with the depth of the potential well being several electron volts (1 eV is approximately $8,000 \text{ cm}^{-1}$ and $1,000 \text{ cm}^{-1}$ is approximately 12 kJ mol^{-1}). This may be contrasted with the curves drawn in the figure.

Although typical in shape the depths of the potential well are now less than 2 cm^{-1} and average separation is over 30 Å (1 Bohr, $a_0 = 0.52917 \text{ Å}$). The

Graham Richards is a lecturer in the Physical Chemistry Laboratory, Oxford University.

curves are those predicted by Stwalley *et al.* for two states of diatomic sodium.

Experimentally these states are visible as angular momentum recoupling extrema in the far wings of resonance lines of Rb and Cs. In particular, the red wing stellite on the 8521 Å line of Cs corresponds to the O_g state illustrated in the figure for Na_2 . The authors hope to prepare molecular beams with significant populations of long-range levels of Li_2 and Na_2 and to observe rotational fine structure in bands involving the long-range states by laser spectroscopy. □

Metallic hydrogen

from N. E. Cussack

A RECENT report by P. S. Hawke and fellow workers at the Lawrence Livermore Laboratory (*Phys. Rev. Lett.* 41, 994; 1978) that highly compressed hydrogen is a metal raises again one of the most interesting questions in the physics of condensed matter, namely, what decides whether a substance is metallic or not? Many theoretical and experimental studies, going back to classical papers by Mott in the 1950s, have made it clear that one decisive parameter is the density. In a way this is obvious; at room temperature liquid mercury, for example, is metallic but its vapour is not. But whether there is a critical density at which a sharp phase change from metal to non-metal occurs and, if so, what electronic mechanism is at work are still open to discussion and much depends on whether the material is crystalline, or structurally disordered as in a fluid.

For crystals, it holds that if the bands of allowed electron energy are either fully occupied or completely empty, then the material is non-metallic. Partially empty bands permit metallic conductivity and therefore, if a full and an empty band can be made by a density change to overlap in energy, they will both become partially full and a metal will be created. The reverse could occur if the density change caused two overlapping energy bands to separate. These are called 'band crossing transitions' and are believed to occur in insulating iodine and metallic ytterbium which, under high pressures, reverse their roles and become a metal and an insulator, respectively. However, other considerations enter for monovalent metals like sodium which have half full energy bands. If these could be reduced in density—not possible in the crystallised

state—it is thought that at a critical density the electrons would condense on to the ions and be bound into neutral centres thus losing their freedom to conduct electricity. Contrary to the band crossing case, such a transition is chiefly an effect of the mutual interactions of the electrons themselves and is known as a Mott transition. There are theoretical reasons for believing that both Mott and band crossing transitions are discontinuous phase changes in crystals.

In fluids, or disordered solids, the whole matter is complicated by the structural chaos and although discontinuous transitions are not ruled out, gradual changes are quite possible. Furthermore, electrons may lose their ability to conduct electricity because the disorder itself confines them to finite regions of the material—an unexpected and subtle modification of the electron wave functions known as 'Anderson localisation'. Fluid mercury reduced in density to about 9 g cm^{-3} at about 1400°C and 1500 bar pressure is believed to become non-metallic by a continuous process involving Anderson localisation.

In all this, the behaviour of hydrogen is particularly challenging for a number of reasons. In the first place, increasing the density of hydrogen sufficiently to give it any chance of becoming metallic requires enormously high pressures—in the megabar range. If the production of such high pressures is accompanied by high temperatures, knowledge of the sample density, or even of whether the hydrogen is liquid or crystalline, will require some understanding of the equation of state. Understanding of both the equation of state and the metal-non-metal transition ought to come relatively easily with hydrogen because it is the simplest element. However, solid hydrogen is a molecular solid and from the point of view of the electron energy band theory has full and empty bands and is therefore an insulator. If it is sufficiently compressed will it undergo a band crossing transition or will it suddenly transform to a monovalent metal with half full bands like the next simplest monovalent metals lithium and sodium? These questions have an interest beyond solid state physics because high pressure hydrogen exists in the larger planets of the Solar System, and planetary magnetic and other properties will be greatly influenced by whether their hydrogen layers are metallic or not.

In their paper, Hawke and collaborators reported high conductivity in a hydrogen sample at a density of 1.06 g cm^{-3} and a pressure of 2 Mbar.

The detection of metallic hydrogen had been reported previously at about 2.8 Mbar by Grigor'ev *et al.* (*JETP Lett.* 16, 201; 1972) but by interpreting an 'anomaly' in the thermodynamic properties rather than by observing high conductivity. These two groups attain the high pressure by causing an explosive to implode a cylindrical container. In the Livermore experiment the imploding steel cylinder compresses magnetic flux which in turn compresses the hydrogen sample inside its silver holder. In such experiments measurements must be carried out by rapid methods during the 10 or 50 microseconds available. Densities, for example, are found by observing the sample volume by X-ray flash radiographs, and conductivities by operating continuously a reflectometer using 15 MHz pulse trains. Other investigations of high pressure hydrogen have tried shock waves and high intensity laser pulses to produce the pressure. It is not impossible to get into the megabar range with static anvil devices and indeed Vereshchagin *et al.* (*JETP Lett.* 21, 85; 1975) claimed in 1975 to have detected metallic hydrogen at 1 Mbar in such a static equipment with the sample at 4.2 K.

There seems little residual doubt that metallic hydrogen can be made and that it needs a small number of megabars to do it. Theoretical estimates of the transition pressure have ranged in the past from about 1 to 20 Mbar and it is of course no easy matter to calculate it. The theoretical approach is to calculate the energies of the atomic metallic and molecular insulating phases separately using solid state methods and then determine the pressure and density at which the free energies of the two phases are equal. One recent calculation (Ramaker *et al.* *Phys. Rev. Lett.* 34, 812; 1975) gave 2.1 Mbar which suggests that theory and experiment are beginning to converge. Hawke *et al.* do not consider that a band crossing transition in the molecular crystal has been ruled out by their work but that an abrupt transition to an atomic phase comparable with metallic sodium is at present favoured.

The interest of this work to physicists of condensed matter is considerable and is enhanced both by a suggestion of Ashcroft's (*Phys. Rev. Lett.* 21, 1748; 1968) that metastable metallic hydrogen (if it exists) might be superconducting at room temperature, and also by the relevance of compressed hydrogen isotopes to nuclear fusion. A growing understanding of this interesting example of a metal-non metal transition can be confidently expected but experiments and theory are both difficult and progress may not be rapid. □

The cladistic revolution—can it make the grade?

from L. B. Halstead

OVER a decade ago a German entomologist, W. Hennig, published his book *Phylogenetic Systematics* (University of Illinois Press, Urbana, 1966) in which he presented a codified methodology for determining the relationships of organisms. As the entire system was designed in such a way that fossils could hardly be included, little serious note was taken of the phenomenon which has become known as cladism. Palaeontologists did not reckon on the ingenuity of certain colleagues, who not only succeeded in applying it to fossil fishes, but managed to convert others to this new system.

It was with this background in mind that a special session was set aside for the topic to be debated at the 26th Symposium of Vertebrate Palaeontology and Comparative Anatomy held in September at the University of Reading.

P. Forey, British Museum (Natural History), outlined the methods and aims of the cladists. Hennig had stressed that it was necessary to distinguish primitive or original characters, which he termed plesiomorphic, from advanced or derived ones, now named apomorphic, in order to ascertain the direction in which evolution had progressed. If a number of lineages can be united on their primitive or plesiomorphic characters the grouping is termed paraphyletic and as such is rejected in the Hennigian scheme, since only strictly monophyletic lineages are recognised. Groups that are united on shared derived characters, or a synapomorphy, can, however, be placed in a valid taxon. The approach of the followers of Hennig is best seen in their construction of cladograms, a system of repeated dichotomies where at each branching a new monophyletic lineage is distinguished by a change in at least one derived character state from the original condition. It is axiomatic that such sister groups are given identical rank. Hence an adjacent lineage and all those subsequently descended from it will have the same taxonomic rank as the preceding lineage. Furthermore there is no place for ancestral species in a cladogram. Where complete sequences from species to species or, as G. MacIntyre (Queen's College, New York) demonstrated, from genus

to genus can be demonstrated, then since they will represent a monophyletic lineage, they must necessarily be included in a single specific taxon.

In order to construct a cladogram it is necessary first to be able to distinguish between original and derived characters and to assemble them in some kind of hierarchy, which as S. Westoll (University of Newcastle) stressed, involved a degree of subjectivity from the very outset.

A cladogram is essentially an hypothesis and its validity is continually tested by the law of parsimony (that no more causes or forces should be assumed than are necessary to account for the facts). An example of this process was illustrated by cladograms presented by P. Janvier (University of Paris) on agnathans and B. Gardiner (Queen Elizabeth College, London) on choanichthyans both of which were demolished within hours of their being proposed. One of the problems of cladism is that fossils can only be considered once the model for their modern representatives had been worked out — a very difficult situation for groups with no agreed living representatives such as heterostracan agnathans and dinosaurs.

Cladism has been extensively applied to insects and the symposium volume edited by Greenwood, Miles and Patterson (*The Interrelationships of fishes*, Academic Press, London, 1973) ranks as a major monument to the application of Hennigian principles to the study of fish.

In contrast the very concept of structural grades is anathema to all cladists. Bonde (*Colloque internat. CNRS* No. 218, 293; 1975) claimed "very little has been achieved in studies of other vertebrate groups, for example by Romer (1966, 1968) and Halstead (1969) where grade classification and paraphyletic groups are repeatedly recommended because of the enormous influence of Simpsonian philosophy." Yet as all teachers and writers concerned with the study of the history of life on this planet well recognise, major evolutionary events can be best understood in terms of major structural grades, which happen to coincide with the major classes of the vertebrates. The idea of separate evolutionary lineages achieving a higher grade of organisation, say from a reptilian to a mammalian grade, should occasion no surprise. In terms of understanding evolution, the fact that many parallel changes can be related to adaptations to similar conditions, is more significant than listing trivial trademarks from which cladograms can be constructed. The aim of the cladists is simply to delineate relationships and to insist on a classification that reflects this in its entirety at the

expense of all other considerations.

Yet if it were possible to trace the ancestry of man back in time to a rhipidistian fish then the entire sequence of stages would be by definition a single monophyletic species. This must surely have happened, yet most zoologists and palaeontologists would surely accept the reasonableness of dividing such a lineage into a number of arbitrary divisions. Commonsense alone would dictate such a course of action. Even though it is possible to take all the end twigs of the evolutionary tree that exist at the present time and construct a hypothetical tree of them, such sharp distinctions must of necessity vanish as they are traced back in time. Evolution is a gradual process that takes place through time and man simply imposes on it his own arbitrary system of classification.

The issue was summed up for many by R. Parrington's (University of Cam-

bridge) exasperated exclamation that according to the cladists a lungfish is more closely related to a cow than to a salmon to which C. Patterson (British Museum (Natural History)) answered "Yes, I cannot see what is wrong in that".

The cladists adhere to the tenets of Hennigism with religious fervour and are already entrenched in some of the major museums in the world. The new evolution and diversity exhibit in the Natural History Museum is destined to be organised to show relationships. If the familiar arrangement of family trees illustrating not only relationships but also their distribution through geological time were to be replaced by elaborate cladograms, I feel it would be a severe error of judgement. Cladograms are difficult enough for experts in the field to comprehend fully, let alone the casual visitor seeking either knowledge or entertainment. □

Earthquakes and fault gouge

from Peter J. Smith

MANY large shallow earthquakes are evidently associated with slip along pre-existing faults in the Earth's crust. But this is not to say that the shallow-earthquake mechanism is anything like fully understood; on the contrary, there are outstanding problems of a quite fundamental nature. For example, laboratory experiments on the frictional sliding of rock carried out during the late 1960s suggested that the shear stress needed to activate rock-on-rock sliding was an order of magnitude higher than that actually present on the San Andreas fault, the latter being deduced from local heat flow studies. In other words, stick-slip seemed the most likely mechanism for shallow earthquakes, but it was by no means clear that, in the San Andreas fault zone at least, the conditions for it actually obtained.

More recently the stress criterion has been relaxed insofar as Hanks (*Pageoph.* 115, 551; 1977) and others have argued for a real stress along the fault rather higher than that previously supposed. Even so, it still seems worthwhile to consider what, if anything, could reduce the shear stress actually required to produce fault slip. An obvious possibility here is the presence of fault gouge, for repeated movements in the highly fractured San Andreas fault

zone are likely to have led to extensive comminution, and the resulting gouge may even have been altered by hydrothermal fluids. Presumably, therefore, this gouge, altered or not and dry or wet, could modify the San Andreas fault's slip behaviour. Indeed, some very recent laboratory experiments (for example, Summers & Byerlee, *Int. J. Rock Mech. Min. Sci.* 14, 155; 1977) have shown that artificial gouge between cut rock surfaces often reduces the coefficient of friction, which makes it all the more surprising that gouge—a long-known substance, after all—has not hitherto played a much greater part in mechanism studies. Yet as Wang *et al.* point out (*Geophys. Res. Lett.* 5, 741; 1978), "Experimental data on gouge under conditions appropriate for upper and mid-crustal environments are still lacking".

But first, does fault gouge actually exist in appreciable quantity at depth in the San Andreas fault zone? Intensely faulted and severely altered rocks combined with clay minerals are certainly present at the surface, even though the lack of detailed mapping and classification makes it difficult to obtain 'representative' samples. Nevertheless, Wang and his colleagues have collected fault gouge from outcrops along both the San Andreas and Hayward faults and, for reason that will soon become clear, have measured P and S seismic wave velocities in them at various confining pressures up to

Peter J. Smith is a Reader in the Department of Earth Sciences at the Open University.

Sicily: the world reservoir for thalassemias and haemoglobinopathies

from Gino Schilirò

SICILY is particularly affected by the hereditary disorders of haemoglobin synthesis (the thalassemias and haemoglobinopathies) that are found all around the Mediterranean. In Sicily the very common β -thalassemias and the many haemoglobinopathies present a serious public health problem. For this reason a Centre for the study and treatment of thalassemias was established in Catania in 1965. This Centre, under the direction of Professor Giuseppe Russo, staffed by physicians, laboratory technicians and nurses, serves not only as a modern treatment facility and follow-up for 250 children with Cooley's anaemia (thalassemia major) but also as a centre of research into thalassemia and of a public education campaign¹. Mass screening to detect thalassemia carriers and those with other haemoglobinopathies, and preventive genetic counselling is an important element of the programme of the Centre. Since the estimated number of Sicilian newborns with Cooley's anaemia is about 105 each year, we are developing a programme for pre-natal diagnosis.

From 1974 to November 1978, 7,143 subjects with suspected hereditary haematological disorders were examined at the Centre: 28.5% of them were carriers of haemoglobin anomalies; 292 were homozygous β -thalassemia, 1,342 β -thalassemia trait and four homozygous $\delta\beta$ -thalassemia. We also found 40 cases of α -thalassemia as α_1 , α_2 and HbH disease, but no cases of hydrops fetalis have been seen. The exceptionally high incidence of the haemoglobinopathies in Sicily is reported in Table 1. The interaction between the thalassemias, with their genetic heterogeneity, and

the haemoglobinopathies gives extremely interesting haematological pictures². If one considers that several abnormal haemoglobins have also been reported by others³ in Sicilians or people of Sicilian ancestry, then it is evident that Sicily, because of its geographical position and history, is a place of unique genetic admixture.

In studies of globin synthesis in randomly chosen patients with Cooley's anaemia, two groups have been distinguished: β^0 -thalassemia (30%) characterised by the absence of β chains, and β^+ -thalassemia (70%) with low levels of β chain synthesis⁴. In collaboration with other Italian and foreign laboratories, research at the Centre has contributed to the clarification of some of the molecular defects of the thalassemias. $\delta\beta$ -Thalassemia has been found to be caused by a genetic deletion⁵. We have also observed that while β globin synthesis is completely absent in cells of patients with β^0 Catania, there is β -globin mRNA in the cytoplasm of the cells of these patients and that the β^0 patients seem to have the normal number of β -globin genes⁶.

Haemoglobin S is the most common abnormal haemoglobin found in Sicily. The mean frequency is 1% although in certain areas it may reach 14%, as in people of African origin. The chemical and physical identity of Hb S in both USA Blacks and Sicilians as well as the presence of

African blood group markers suggest that Hb S may have spread from Africa to Sicily⁷. Another haemoglobin we have found is Hb C, of known African origin. Among the 18 families with Hb J Oxford reported in the literature, 13 are Sicilian or of Sicilian ancestry. Carriers of this haemoglobin are asymptomatic and discovered only by screening. It should be noted that in all our cases the rate of Hb J Oxford was about 25%, supporting the hypothesis that α chain production is controlled by four genes. A case of interaction between homozygous β^0 -thalassemia and Hb J Oxford, producing a new alkali-resistant haemoglobin $\alpha_2^J \gamma_2^F$, is an example of the wide range of possible genetic combinations in this island with so many disorders of haemoglobin synthesis⁸.

Although these data come from a selected population at risk, the phenomenon still remains extremely important. In fact, since 1893 many people have emigrated from Sicily to North and South America, to Australia and, more recently, to North Europe. Especially if one considers that, in the 20 years between 1951-71, 1,010,963 Sicilians emigrated, the Sicilian haemoglobin abnormalities become of general interest.

G. Schilirò is in the Department of Pediatrics, at the University of Catania, Italy.

Abnormal haemoglobins found among 7,143 blood specimens

Genotype	Number	Structural* alteration	Genetic combinations found (number of cases)
Hb S	277	β^0 Glu $\rightarrow$ Val	S-S(8) S-A(194) S- β^0 Th(30) S- β^+ Th(55)
J-Oxford	20	α^{15} Gly $\rightarrow$ Asp	J- β^0 Th(1) A-J(19)
Lepore Boston	17	α_2 ($\delta^{87} \beta^{116}$) ₂	I- β Th (5)
Hb C	10	β^0 Glu $\rightarrow$ Igs	A-C (8) C- β^+ Th (1) (C- β^0 Th) (1)
J-Baltimore	9	β^{16} Gly $\rightarrow$ Asp	J- β^+ Th (1) A-J (8)
G-San Josè	3	β^7 Glu $\rightarrow$ Gly	G- β^0 Th (1) G- α_1 Th (2)
D-Punjab	3	β^{121} Glu $\rightarrow$ Gln	D-A (3)
Shepherd's Bush	2	β^{74} Gly $\rightarrow$ Asp	S B-A (2)
G-Copenhagen	1	β^{47} Asp $\rightarrow$ Asn	G-A (1)
Incompletely characterised	2		

*Characterisation of the abnormal haemoglobins was performed by L. Tentori and co-workers at the Centro di Riferimento per le Talassemie ed Emoglobinopatie in Rome.

- Schilirò et al. *Lancet* i, 598 (1976).
- Russo et al. *Acta haemat.* 28, 329 (1962); *Blood* 42, 763 (1973); *Pediatr. Res.* 11, 1019 (1978); Musumeci et al. *J. med. Genet.* (in the press).
- See for example Sansone et al. *Acta haemat.* 43, 40 (1970); Riccio et al. *FEBS Lett.* 39, 700 (1974); Sharma et al. *Biochim. biophys. Acta* 393, 379 (1975).
- Schilirò et al. *Acta haemat.* 60, 193 (1978).
- Ottolenghi et al. *Cell* 9, 71 (1976).
- Ramirez et al. *Clin. Res.* 24, 442A (1976); *Nature* 263, 471 (1976).
- Sandler et al. *Acta haemat.* (in the press).
- Schilirò et al. *Blood* 48, 369 (1976).

2 kbar (equivalent to a depth in the Earth of 8-10 km).

The results are simply stated. The P wave velocities in the fault gouge samples and their variation with pressure (depth) are very similar to those obtained by Healey and Peake (*Bull. Seismol. Soc. Amer.* 65, 1177; 1975) for

the central Californian region of the crust bounded by the San Andreas and the San Benito faults. Moreover, both the P and S wave velocities in the gouge and the P wave velocities obtained by Healey and Peake are significantly lower than those in intact rocks, whether the intact rocks be in

situ crust in the region outside the San Andreas fault zone or samples brought from the San Andreas fault zone itself (Lin, thesis, University of California, 1977). In other words, the low seismic velocities in the San Andreas fault zone are consistent with the presence of appreciable amounts of gouge to depths

of at least 8-10 km. Support for this interpretation also comes from gravity data, for to explain the steep gradient in the Bouguer gravity profile near the San Andreas fault zone it is apparently necessary to invoke the presence of a slab of low density ($\sim 2.6 \text{ g cm}^{-3}$) extending to a depth of 8-10 km. A density of 2.6 g cm^{-3} is typical of fault gouge.

Wang and his coworkers go no further with their experiments; but they have gone far enough to show that fault gouge as a possible modifier must not be ignored when it comes to considering the causes of shallow earthquakes. $\square$

Phase transitions in spin glasses

from a Correspondent

If the temperature of a supercooled liquid is lowered sufficiently, it may form a glass. For twenty years or more there has been an inconclusive debate as to whether this change is of purely kinetic origin, occurring when the relaxation times of the liquid become longer than the patience of the experimenter, whereupon the liquid appears to have solid properties, or whether the change reflects an underlying phase transition. It looks as though the debate on the causes of the similar phenomena found in spin glasses may be just as lengthy and inconclusive.

A spin glass is typically a substitutional alloy of a few percentage of a transition metal such as iron or manganese in a host metal such as copper or gold. The transition metal impurities—the spins of the spin glass—can be taken as being approximately randomly distributed throughout the alloy. They interact with each other through the conduction electrons of the host metal to give a long-range coupling between the spins whose sign depends on the distance between the spins. This means that the interaction between some spins is ferromagnetic in sign which favours their parallel alignment while that between other pairs of spins is antiferromagnetic which favours their antiparallel alignment. The net effect of such a mixture of competing interactions is to produce a ground state in which the orientation of the spins is chaotic and which has no resultant magnetic moment.

When cooled below its freezing temperature T_F , the properties of a spin glass are radically altered in much the same way as the properties of a supercooled liquid are altered below the glass transition temperature T_g . Above T_F an induced magnetic moment rapidly decays away on a microscopic time

scale when the magnetic field which induced it is switched off. Below T_F an induced moment can take hours to decay away. One can immediately deduce that some relaxation processes in a spin glass take place very slowly. Another characteristic is the 'cusp' in the magnetic susceptibility at the freezing temperature. The susceptibility as a function of temperature rises at T_F to a maximum which is quite sharp, especially when the susceptibility is measured using small applied fields.

Non-analytic behaviour such as a cusp usually indicates that a phase transition must be taking place. In 1975, Edwards and Anderson (*J. Phys. F6*, 1927; 1975) proposed a phase transition theory for spin glasses which gave a seemingly satisfactory explanation of the cusp. The order parameter q of their theory is the probability that any given spin is still pointing after an infinitely long time in the same direction as it was at some initial time. Above T_F , q is zero, but below T_F , which is identified as the phase transition temperature, it is non-zero and increases to unity as the temperature approaches absolute zero.

Although we live in a three-dimensional world, it is both interesting and useful to examine (theoretically!) the nature of a phase transition in worlds of different dimensionality. There is always a 'lower critical dimension' below which the phase transition will not take place. Thus the Heisenberg type of ferromagnet does not have a phase transition in two dimensions for its lower critical dimension is two. The lower critical dimension for the spin glass transition is controversial, with arguments being advanced for two, three and four dimensions. Numerical evidence derived from high-temperature series expansions seems to favour four (Reed, *Phys. Lett. A68*, 473; 1978). If that really is the case, then the phase transition theory of spin glass behaviour fails and an alternate explanation must be sought.

An explanation which actually pre-dates the phase transition theory is that spin glass behaviour is entirely a non-equilibrium effect resulting from the long relaxation times which exist within spin glasses at low temperatures. The relaxation modes probably involve activation processes in which a potential barrier separating two easy orientations of a spin has to be surmounted. This will produce relaxation times which follow an Arrhenius law and so rapidly increase at low temperatures. The 'cusp' in the susceptibility is explained on this hypothesis by observing that at temperatures below T_F only a fraction of the total spins present will contribute to the susceptibility, namely those spins or groups of spins which can relax on a time scale shorter than

the time scale of the experiment. The other spins are effectively frozen in fixed orientations and do not contribute. For a given experimental time scale, the further the temperature is lowered below T_F the larger the fraction of frozen spins becomes and hence the smaller will be the observed susceptibility. A corollary of this hypothesis is that the freezing temperature T_F should alter with the experimental time scale. Lengthening the time scale should enable more spins to contribute to the susceptibility and so depress the apparent freezing temperature. A similar effect occurs at the glass transition in ordinary glasses where T_g falls when the liquid is allowed a longer time to come to equilibrium.

A recent experiment of Murani and Heidmann (*Phys. Rev. Lett. 41*, 1402; 1978) clearly indicates that lengthening the time scale depresses T_F . They have carried out neutron inelastic scattering measurements on a Cu-8 at % Mn alloy at three different energy resolutions. The temperature dependence of the elastic magnetic cross section can be extracted from the data. It is related to the Edwards-Anderson order parameter q on a phase transition hypothesis and to the number of frozen spins on a relaxation time hypothesis. The elastic cross section increases below a certain temperature but the temperature at which this happens, the nominal freezing temperature, varies with the energy resolution employed. They argue that the time scale in their experiment is proportional to the inverse of the energy resolution. For a time scale of 10^{-11} s, T_F is 75 K, whereas T_F is about 40 K on a time scale of 10^{-2} s (this latter data is from separate experiments on the a.c. susceptibility). At first sight this appears to be strong evidence against a phase transition (since a phase transition should take place at a unique temperature) and in good accord with the explanation of spin glass behaviour in terms of long relaxation times and frozen out spins.

Unfortunately it is not as simple as this. A phase transition could be accompanied by long relaxation times which would produce an apparent freezing temperature that decreased with increasing experimental time scale, but which eventually saturates at the true transition temperature T_F for a hypothetical experiment with an infinitely long time scale. There is just a hint of such a possibility within their data but it is not sufficiently strong to be compelling evidence. It is clearly going to be very difficult to decide from experiments of this kind whether spin glass behaviour is or is not due to a phase transition. A new idea is needed for a conclusive experimental test to distinguish the rival approaches.

review article

Electron cooling and its applications in elementary particle physics

N. S. Dikansky, I. N. Meshkov & A. N. Skrinsky

Institute of Nuclear Physics, Siberian Branch, Academy of Sciences of the USSR, Novosibirsk 90, USSR

In recent years, experiments in high-energy physics have shown that one of the most promising means of studying the properties of elementary particles is by colliding beams. First, they open up a region of interaction energies at present unattainable using conventional accelerators. Second, they allow experiments to be performed in extremely clean conditions, when the observed interaction between two colliding particles is not contaminated by the presence of secondary interactions as in the conventional particle accelerator-fixed target scheme. This article describes the electron cooling method applied to proton-antiproton colliding beams.

THE first experiments carried out with electron-electron colliding beams in 1965 at Stanford and Novosibirsk gave a convincing demonstration of the potential of this technique in high-energy physics. In these experiments, from practical considerations, electrons were preferred, however, to obtain more information on elementary particles, colliding beams of particles and anti-particles are of much greater interest. In the interaction between a particle and an anti-particle, a state is produced in which the total charges of all kinds, such as electrical, lepton, and baryon, equal zero, and the system is then released from the conditions imposed by the conservation laws of these charges. In fact, this state is equivalent to a pure energy-blob, capable of creating any particle-antiparticle pairs permitted by the energy-momentum conservation law. The first experiments on colliding beams with particle-antiparticle (electron-positron) annihilation were carried out at the Nuclear Physics Institute in Novosibirsk on the VEPP-2 apparatus (abbreviations of Vstrechnye Elektron-Pozitronnye Puchki—electron-positron colliding beams). The energy of the particles in these experiments reached 0.67 GeV. There are six electron-positron colliding-beam devices at various accelerator centres throughout the world (Novosibirsk, Stanford, Orsay, Hamburg, Frascati), with energy ranges from 2×0.2 to 2×4 GeV. Three more electron-positron colliding beam machines are under construction (Novosibirsk, Hamburg and Stanford) with an energy range of 2×4 to 2×20 GeV.

Experimental constraints

Electrons and their antiparticles are interesting as they represent 'point' objects: according to current experimental data the size of these particles does not exceed 10^{-15} cm. Unfortunately, for electrons and positrons the practical energy limit for colliding beams in their traditional configuration is not very far from the level attained today. This limit is associated with synchrotron radiation, the intensity of which increases as the fourth power of the energy of the particle, and decreases only as the first power of the orbit radius, so that increasing the

size of the ring provides no real solution to the problem. Colliding beams of heavy particles do not have this problem as the intensity of synchrotron radiation is inversely proportional to the fourth power of the rest mass of the particle. But here, unhampered by energy losses, we have at the same time lost a mechanism which quenches the oscillations of the particles in the storage ring and so permits multiple storage of particles in the same phase volume.

For proton-proton colliding beams this does not in principle create any constraints as there are ways of producing intense beams of protons. The first proton intersecting storage rings have been in operation at CERN since 1971. However, for antiparticle beams accumulation is essential because of the low production yield. Antiparticles are generated with a large phase volume and it is therefore important to incorporate a mechanism for reducing the phase volume of the beam.

Such a mechanism was found by Budker, for heavy particles, who proposed the electron cooling method. Subsequently, it was realised that this method permits the accumulation of large antiproton currents and allows experiments to be carried out on proton-antiproton ($p\bar{p}$) collisions (A. Skrinsky). Hence, in 1966 at the Paris conference on colliding beams¹, in addition to the electron cooling method, Budker presented a $p\bar{p}$ colliding beam project worked out at our Institute. Although both the electron cooling method and the $p\bar{p}$ -colliding beam project evoked great interest little progress was made, for the following reasons. First, the electron cooling method had not been demonstrated experimentally. It seemed to be an exotic idea with little hope of realisation. Second, during the 1960s, experiments on $p\bar{p}$ colliding beams were considered an important but extremely complicated way of supplementing $p\bar{p}$ experiments studying the nature of strong interactions.

Today the situation has changed radically. The observation of point-like constituents in protons which are now identified with quarks, has forced us to reconsider our attitudes to $p\bar{p}$ colliding beams: from the point of view of the quark model of strongly interacting particles, such beams are equivalent to colliding

beams of quarks and antiquarks. This implies that $p\bar{p}$ colliding beams provide fundamental information similar to that obtained from electron-positron (e^-e^+) colliding beams. During the interactions, energy is liberated 'at a point' that is it is not smeared over an extended object; which allows information to be obtained on the properties of matter at limitingly small distances. According to current ideas, the proton (antiproton) contains three quarks, so that $p\bar{p}$ -colliding beams with a given energy are equivalent to e^-e^+ colliding beams with energy three times less.

Finally, during the past few years, a decisive step forward has been made towards the creation of $p\bar{p}$ -colliding beams: the electron cooling method has been studied and developed experimentally. This article describes this method and its potential for elementary particle and nuclear physics.

Machine luminosity

The efficiency of a colliding beam machine is characterised by the luminosity L , a quantity equal to the number of interactions per unit time during the collision of two beams per unit interaction cross-section. Then the counting rate in the i -th reaction channel with cross-section σ_i is

$$\frac{dN_i}{dt} = L\sigma_i \quad (1)$$

In the case of two bunches with N_+ and N_- colliding particles the luminosity is given by the relationship

$$L = \frac{N_+ N_-}{S} f \quad (2)$$

where f is the frequency of revolution of the particles in the ring, and S is the cross-section of the larger of the bunches. The fundamental problem with colliding beams is to obtain the maximum possible luminosity. We note that equation (2) holds if the detector sees the whole intersecting region. Hence it is advantageous to work with beams of short bunches.

When working with antiparticle beams the problem of intensity is especially serious. Antiparticles are produced by accelerators in comparatively small quantities, being ejected from a target by a beam of particles of sufficiently high energy. Moreover only a small fraction of the antiparticles thus generated can be injected into the storage ring: the permissible energy spread in a beam of anti-particles is usually a fraction of a per cent of the mean energy of the injected particles. For example, in the VEPP-2M storage ring out of 10^{11} electrons it was possible to obtain only of the order of 10^7 useful positrons, while for experiments with colliding beams it is necessary to have at least 10^{10} – 10^{11} particles. Thus the only course is the accumulation of antiparticles: by multiple injection into the same phase volume of the storage ring. For this, it is necessary to constrict the phase volume of the injected beam before the next injected pulse, thus releasing phase space for the next lot of particles. This increase in the density of the beam in phase space takes place if friction is introduced into the system of particles.

Particle oscillations

A particle orbiting in a storage ring is known to perform oscillations in the focusing fields of the ring magnets with respect to some equilibrium trajectory. Three-dimensional oscillations in the (x, y, z) coordinate space correspond to motion over the surface of a six-dimensional ellipsoid in the coordinate-momentum (p_x, p_y, p_z, x, y, z) phase space. The set of N particles occupies some volume in phase space which remains invariant with time if the oscillations are not damped. If, however, a frictional force acts on the particles, the amplitude of their oscillations decreases with time, and, consequently, the volume which they occupy decreases.

If the frictional force $\mathbf{F}$ is proportional to the total velocity $\mathbf{V}$ of the particle

$$\mathbf{F} = -\lambda m \mathbf{V} \quad (3)$$

the decrease of the amplitude (A) of oscillations in the case of weak friction is exponential ($\lambda \ll \omega$), where m and ω are the mass and frequency of the oscillations of the particle, with a decrement

$$\lambda = \frac{-1}{A} \frac{dA}{dt} \quad (4)$$

In addition to friction, diffusion processes are always present in a real system, the action of these may be characterised by the diffusion coefficient D :

$$\frac{da^2}{dt} = D \quad (5)$$

Then the damping of the oscillations and the constriction of the phase volume reaches some steady equilibrium value when friction and diffusion are equal:

$$a_s^2 = \frac{2D}{\lambda} \quad (6)$$

These equations define two important characteristics of the storage ring: the rate of compression of the phase volume λ and the minimum (steady) dimension a_s of the stored beam, on which the luminosity of the storage ring depends.

Radiation friction

An example of such a dissipative process is the radiation friction for electrons and positrons due to synchrotron radiation which we have already mentioned. The force of reaction of the radiation is directed opposite to the total velocity of the particle. If the accelerating system of the storage ring keeps the longitudinal component of momentum constant on average, then the radiation friction quenches the deviations of the momentum from the equilibrium value. Radiation friction is a powerful dissipative process: typical damping times of the electron and positron oscillations in operational accelerators are from 1 s down to milliseconds. This allows the problem of multiple stacking to be solved. Hence electron-positron colliding beams have high luminosities. This mechanism, as noted above, does not apply to heavy particles—protons, antiprotons or ions.

Electron friction

In Budker's method, the friction in a beam of heavy particles circulating in a storage ring is introduced by the flow of electrons. For this, it is necessary to arrange in the straight section of the trajectory of the heavy particles, their interaction with electrons of the same mean velocity as the heavy particles. A particle whose velocity differs from the mean experiences, as a result of collisions with the electrons, the action of a frictional force which tends to damp its oscillations. In the simplest case of homogeneous electron beam with maxwellian velocity distribution the damping for each of the three spatial degrees of freedom is identical:

$$\tau = \frac{1}{\lambda} \approx \frac{0.5}{r_e r_p c L} \frac{\gamma^2}{n_e \eta} \left[\left(\frac{T_e}{mc^2} \right)^{3/2} + \left(\frac{T_p}{Mc^2} \right)^{3/2} \right] \quad (7)$$

Here r_e, r_p are the classical radii of the electron and the heavy particle, m, M, T_e, T_p are their masses and temperatures, n_e is the electron density in the cooling current, $\gamma = (1 - \beta^2)^{-1/2}$ is the relativistic factor, β is the mean velocity of the particles in units of the velocity of light c and η is the fraction of the perimeter of the orbit occupied by the cooling section.

This model is equivalent to a homogeneous two-component plasma (particles–electrons), while equation (7) is the relaxation

time of the plasma (equalisation time of the temperature of both components). To find the steady amplitudes of oscillations, equation (6), in this simple case there is no need to calculate the diffusion coefficient for a heavy particle in the electron gas: it is clear that the steady state corresponds to the equilibrium temperature of both components:

$$T_p = T_e, \text{ or } V_p = \left(\frac{m}{M}\right)^{1/2} V_e \quad (8)$$

Thus if the temperature of the cooling electrons is low ($T_e \ll T_p$), electron cooling works effectively. (All units are given in the centre mass rest frame.)

Naturally, the kinematics of a real cooling process are more complicated. The detailed motion of the proton beam in the storage ring must be taken into account, and also the effect of errors in the equalisation of the mean velocities of electrons and protons, coherent fluctuations of the electron beam, and so on.

Electron beam cooling

In the experiments described below, various effects associated with special features of the electron beam have been studied: the role of the sharp anisotropy of the velocity distribution function of the electrons in a beam accelerated by an electrostatic field², and the effect of a strong magnetic field on the cooling efficiency⁴. Fortunately, the latter effects increase the cooling properties of the electron beam. Today the process of electron

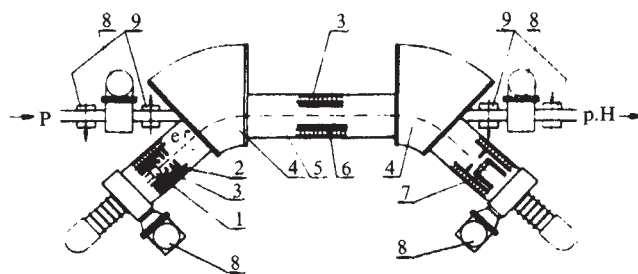


Fig. 2 Layout of an electron-beam installation: 1, electron gun; 2, anodes; 3, solenoid; 4, bending section of electron beam; 5, cooling section; 6, vacuum chamber; 7, collector; 8, vacuum pumps; 9, correction magnets.

effective length of the cooling section is only 1/50 of the perimeter ($l = 1 \text{ m}$, $\eta = 2 \times 10^{-2}$). An intense electron beam with low transverse velocities is obtained from a special electron gun, it is then guided in a longitudinal homogeneous magnetic field of 1 kG which extends up to the collector. The electron energy is recovered in the apparatus; this is especially important for high-energy electron-cooling devices.

The effect of electron cooling may be observed immediately, using a thin (0.5 mm) jet of magnesium vapour to measure the density of the proton beam. The jet intersects the proton beam, and the secondary electrons are collected by the electrostatic field onto a scintillation screen and recorded by a photomultiplier. When the jet scans the storage ring aperture, the transverse density distribution of the proton beam can be reproduced on an oscilloscope screen.

Without the electron beam, an increase in the dimension of the proton beam is observed due to multiple scattering on the residual gas. After the electron beam is switched in, the dimension of the proton beam decreases sharply. The effect of damping is seen especially clearly if the jet is kept at the centre of the damped proton beam and its dimension is increased by means of an inflector pulse. Then the density at the centre of the proton beam which is reduced after the pulse will increase to its equilibrium value within a characteristic time, τ .

The steady dimension of the proton beam is especially easy to measure with high accuracy by observing the beam of fast hydrogen atoms which is created in the cooling section as a result of the recombination of protons with the cooling electrons. It is interesting that even without special efforts more than 10% of the protons are converted into fast neutral atoms during the lifetime of the beam. At a distance of 10 m from the cooling section the dimension of the neutral beam is only 0.7 mm.

The damping of the energy spread of the proton beam is well demonstrated when the RF system of the storage ring is disconnected. If the energy of the electron beam is varied slightly, then the energy of the proton beam also varies and the radius of the proton orbit changes accordingly. If the energy of the electron beam and the magnetic field of the storage ring are varied simultaneously, then the proton beam can be accelerated or retarded without changing the equilibrium position of the orbit. Thus, for example, a proton beam was accelerated from 65 to 85 MeV during 200 s, corresponding to an energy gain by the protons of 0.3 eV per revolution.

Electron cooling was successfully verified in the proton energy range from 1.5 MeV (injection energy) to 85 MeV (0.4 GeV/c momentum) both for a freely circulating and for a bunched beam. The results obtained so far and the corresponding experimental conditions are given in Table 1.

The values of the cooling time τ and the strength of the longitudinal frictional force obtained in the experiment are better by almost two orders of magnitude than the values expected from the preliminary estimates¹. It is now understood that cooling really depends on many different factors, the relative role of which is determined by the actual conditions in the storage ring with electron cooling.

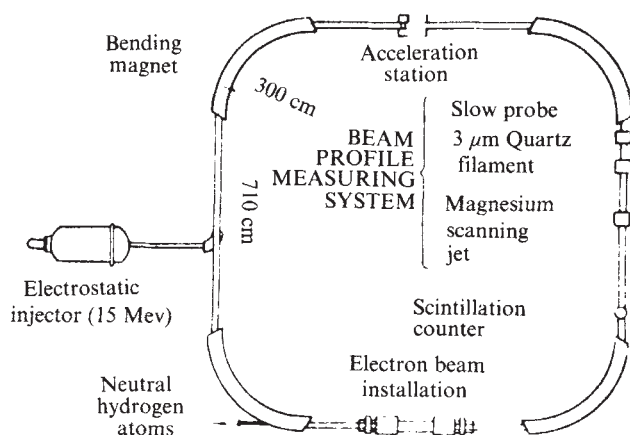


Fig. 1 Layout of the NAP-M proton accelerator.

cooling is fairly well understood and has been studied extensively.

The practical realisation of this technique started at the Nuclear Physics Institute in 1967. In 1974, the first experimental results on the electron cooling of protons were obtained, confirming that the fundamental ideas were correct (G. I. Budker, Ya. S. Derbenev, N. S. Dukansky, I. N. Meshkov, V. V. Parkhomchuk, D. V. Pestrikov, R. A. Salimov, A. N. Skrinky, B. N. Sukhina), while further studies of the technique by the same group indicated its high efficiency.

The experiments were carried out using a facility called the NAP-M (NAP is nakopitel antiprotonov—antiproton storage ring; M = model). The apparatus is shown in Fig. 1. Its perimeter is 47m and the length of each of the four straight sections is 7.2 m. The mean pressure of the residual gas in the vacuum chamber of NAP-M with the electron beam is $< 5 \times 10^{-10}$ torr. The stability of the magnetic field of the storage ring in the electron cooling regime is better than 1×10^{-5} . The control of the ring and the analysis of the experimental data was carried out by an on-line computer.

The electron beam for the cooling process comes from a source in the straight section of the storage ring (Fig. 2). The

Table 1 Typical experimental conditions and best results

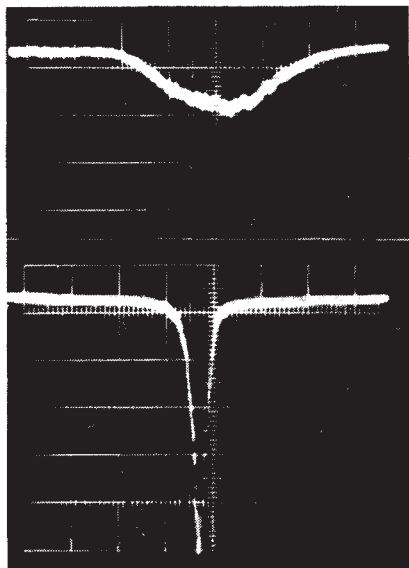
Proton momentum	0.35 GeV/c
Proton current	100 μ A
Electron energy	35 keV
Electron current	0.5 A
η	2×10^{-2}
Effective transverse electron temperature	0.2 eV
Effective longitudinal electron temperature	$< 10^{-4}$ eV
Longitudinal magnetic field	1 kG
Cooling time	50 ms
Maximum longitudinal friction	1.5 MeV s^{-1}
Steady parameters of the cooled proton beam	
Diameter	0.5 mm
Angular spread	$\pm 5 \times 10^{-5}$
Momentum spread	$< 1 \times 10^{-5}$
Energy spread	$< 2 \text{ keV}$
Lifetime	10 h

Experimental possibilities

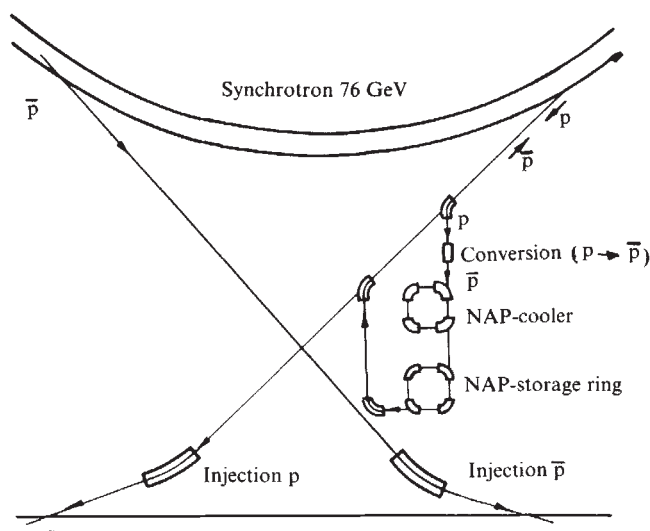
Electron cooling opens up possibilities in various areas of elementary particle and nuclear physics. Our survey⁴ is devoted to their detailed discussion but here we shall confine ourselves to the most interesting examples which illustrate the promising nature of the electron cooling method.

Many experimental studies concern the storage of antiprotons, which is probably the most basic application of electron cooling. As an example, we shall give the parameters of the antiproton storage system proposed by the Nuclear Physics Institute for a planned accelerator storage ring complex at Serpukhov⁴. A simplified scheme for the complex is given in Fig. 4. Using the Serpukhov accelerator with the booster at present under construction (10^{13} protons per pulse, $2 \times 10^{12} \text{ p s}^{-1}$), it is possible to accumulate $0.5 \times 10^6 \text{ p s}^{-1}$ of 1.8 GeV/c momentum even without phase constriction of the proton beam in the accelerator. If we assume, optimistically, that it is possible to use the designed intensity of a proton beam of 10^{14} protons per pulse, then it is possible for a certain configuration of the system⁴ to obtain $1 \times 10^9 \text{ p s}^{-1}$. But even an intensity of $0.5 \times 10^6 \text{ p s}^{-1}$ would give a luminosity of up to $10^{31} \text{ cm}^{-2} \text{ s}^{-1}$ for $p\bar{p}$ intersecting beams of an energy of $2 \times 10^3 \text{ GeV}$. The first achievements in the storage of antiprotons have already allowed a wide range of new experiments to be carried out.

Fig. 3 Density distribution in proton beam before cooling (upper oscillogram) and after cooling (lower oscillogram). Horizontal scale: one large division corresponds to 1 mm.



In electron-cooled antiproton storage rings it is possible to perform a wide range of experiments with 'super-thin' internal gas and vapour targets. Here all the reaction products will be available for detection and analysis, including the slowly ionising fragments. In the superthin target case the antiproton energy losses and their fluctuation together with the multiple scattering of antiprotons on the target atoms are suppressed by the electron cooling, and the life-time of the antiprotons is determined only by interactions with the target. For a good vacuum, the scattering on the residual gas is negligibly small. At high energies and for an accumulation rate of 10^8 p s^{-1} , the luminosity may reach $10^{33} \text{ cm}^{-2} \text{ s}^{-1}$, which is sufficient for a wide range of experiments.



Straight section of accelerator-storage ring 1 TeV

Fig. 4 Layout of proton-antiproton facility.

Superthin internal targets are particularly suitable for precision spectrometer experiments. Electron cooling in antiproton-proton and antiproton-nucleus experiments gives an energy resolution of the order of 10^{-4} – 10^{-6} with satisfactory luminosity. Similar experiments with proton beams give the same high resolution, which is comparable with what can be attained in tandem accelerators and meson factories.

An important factor in spectrometer experiments is the possibility of very precise measurements of the absolute energy of particles of known mass in an accelerator by measuring the voltage accelerating the cooling electrons. Even today an accuracy of better than 10^{-5} may be attained.

Special possibilities are opening up in the study of $p\bar{p}$ interactions in experiments with continuously cooled beams in the energy range of 10 GeV per beam. Simultaneous cooling of the intersecting beams is possible even in a single ring (because the electrons, which cool protons, for example, will not perturb the antiproton beam, as its velocity has opposite direction). This set-up in comparison with the superthin target variant, may prove interesting if it is required to eliminate unwanted interactions with the target electrons and to reduce beam dispersion.

In particular, it is possible to investigate ψ -meson with a resolution much better than their intrinsic width, which is unattainable even in current experiments on e^+e^- intersecting beams. Of course, only a very small proportion of all $p\bar{p}$ annihilation will take place via ψ -mesons even at resonance. Hence it is natural to look for the leptonic modes

$$p + \bar{p} \rightarrow \psi \rightarrow \begin{cases} e^+ + e^- \\ \mu^+ + \mu^- \end{cases}$$

The cross-sections of such processes at the resonance maximum is of the order of 100 nb. In a special storage ring with strong

focusing at the intersection regions it is possible to obtain tens of thousands of such events per day.

We note that in $p\bar{p}$ collisions it is possible to obtain directly narrow states of the ψ -meson type with less strict constraints on quantum numbers than with e^-e^+ colliding beams, where the quantum numbers of such a state must be equal to those of the photon. It will be particularly easy to distinguish states with significant non-hadronic decay modes. An example of this type of meson is the $\chi(2750)$, decaying into two photons.

Using stored antiprotons, it is possible to obtain a beam of antihydrogen atoms. For this a parallel beam of cold positrons with the same mean velocity must be injected into one of the storage ring straight sections, through the electron-cooled antiproton beam. As a result of radiative recombination, the bulk of the antiprotons are converted into antihydrogen. Even on the NAP-M without special effort 10% of the protons are transformed into a beam of fast atoms.

Using this method it is possible to obtain beams of antiatoms of very low energy, which may permit neutral antihydrogen to be accumulated in a magnetic system which acts on antiatoms through the magnetic moment of the positron.

An important aspect of electron-cooled antiproton storage rings is the production of polarised beams of antiprotons. Today we consider the main variant to be the interaction of the stored antiproton beam with polarised superthin atomic hydrogen targets. According to existing data it is still difficult to choose optimal conditions combining a good degree of polarisation with minimal losses of antiproton intensity.

One possible scheme is the interaction with longitudinally polarised protons in the storage ring section where the longitudinal polarisation of the antiprotons is stable. Here the difference in the total interaction cross-sections of antiprotons and protons of different relative helicities is used.

Polarisation experiments in storage rings seem feasible with antiproton beams. The method is adequate too for experiments with polarised protons and deuterons. Of course, polarised beams can only be obtained by directly storing polarised particles. The beams of accelerated polarised protons obtained so far and the existing polarised gas targets yield luminosities of not less than $10^{30} \text{ cm}^{-2} \text{ s}^{-1}$ under very clean conditions.

Electron cooling machines could already be proposed which may allow the intensity of the accelerated polarised proton beams to be increased to the level of existing proton beams of high-energy accelerators.

The use of electron cooling opens up the possibility of storing pure beams of antideuterons. The rate of accumulation of such beams is 10^{-4} of the rate of accumulation of antiprotons. In the superthin deuteron target case this would give luminosities of up to $10^{29} \text{ cm}^{-2} \text{ s}^{-1}$, which makes it possible to study in detail neutron-antineutron strong interactions, including annihilation processes. To distinguish these possibilities, it is necessary to register the antiprotons with small momentum loss and to select events with zero total charge of the remaining final particles, or to register the low-energy proton remaining after the decay of the neutron. The storage of antideuterons would be accompanied by the capture of mesons, antiprotons and antihyperons of momenta corresponding to the field and radius of the storage ring. It is not difficult to obtain a clean anti-deuteron beam: the mesons and antihyperons will decay rapidly, and the antiprotons may be removed by accelerating or retarding the antideuterons after storage with the aid of an electron beam, as described above.

There are also interesting prospects for experiments with stored ions, first in the superthin target regime using high monochromaticity of the cooled beams. Especially important is the use of electron cooling with ions which are difficult to obtain, principally heavy ions. However, working with these is complicated by radiative recombination with cooling electrons (the cooling is weakened in comparison with the radiative recombination proportionally to the mass of the ions). Hence it may be possible to cool these ions too, but it is impossible to keep them cooled for a long time. This problem may be solved by using

proton cooling, for which the ion storage ring is supplemented by a proton storage ring, the straight section being common to both rings. The low temperature of the protons must be maintained by electron cooling.

Nevertheless, the most exciting of the immediate applications of electron cooling is still its use for experiments with intersecting $p\bar{p}$ beams at high energies.

By the time the first report on the principle of electron cooling appeared, a proposal for $p\bar{p}$ colliding beams (VAPP) had already been worked out in Novosibirsk. This proposal was published in 1971. It included a main storage ring with an energy of the colliding particles $2 \times 23 \text{ GeV}$, the NAP antiproton storage ring of 1.1 GeV energy and an antiproton source. More recently a Fermilab project proposes to use the 500 GeV synchrotron as a $p\bar{p}$ colliding beam ring. Somewhat earlier the Nuclear Physics Institute had proposed the $2 \times 2,000 \text{ GeV}$ $p\bar{p}$ colliding beam project for UNK mentioned above. In this conservative approach a luminosity of $3 \times 10^{31} \text{ cm}^{-2} \text{ s}^{-1}$ may be attained, which is more than sufficient for carrying out first generation experiments.

The Fermilab project is in principle analogous to the VAPP/UNK project, except for one important point: the antiprotons 'generated' with an optimal energy of around 4 GeV after cooling are retarded to an energy close to 200 MeV which allows a certain gain in the storage rate. A synchrotron designed for periodic operation seems to allow intersecting beams of $2 \times 300 \text{ GeV}$ energy.

At CERN a project has been approved for a $p\bar{p}$ storage ring based on the 'stochastic cooling' method proposed in 1970 by Van der Meer.

Proton-antiproton ('quark-antiquark') intersecting beams permit the study of lepton pair creation processes.

$$p\bar{p} \rightarrow x \rightarrow \begin{cases} e^+ + e^- \\ \mu^+ + \mu^- \end{cases}$$

in which both the electromagnetic and the weak interactions have a role. In particular, if the neutral intermediate boson Z^0 exists it may be found in these interactions if the total energy of the proton and antiproton exceeds the Z^0 rest mass by a factor of 3–6. Present estimates for its production require proton-antiproton beam energies in the region of $2 \times (150\text{--}200) \text{ GeV}$. In this energy range, it would also be possible to observe the proposed charged vector bosons $W^\pm$ of the weak interaction. It is also possible to search for heavy leptons in such experiments.

One can also study reactions in which a quark-antiquark pair is annihilated into a neutrino-antineutrino pair. Such a reaction is characterised by a visible energy reduced by a considerable fraction of the initial energy of the $p\bar{p}$ pair (of the order of 1/3); this loss will take place almost symmetrically with respect to the initial particles.

More detailed investigation will be possible with polarised $p\bar{p}$, and also pp and deuteron colliding beams. At the high energies considered, it is especially important to carry out experiments with longitudinally polarised particles at the intersection region. Electron cooling allows us to solve the problem of producing polarised proton and antiproton beams of the required intensity.

The quark approach to the interaction of protons and antiprotons emphasises the highly promising nature of experiments on colliding $p\bar{p}$ beams. But this does not exclude the possibility that the real physics of the interaction of matter and antimatter at super-high energies will prove to be still more interesting and entirely unexpected. In any case, there is no doubt that the first answer to these questions will be obtained in the next few years.

1. Budker, G. I. *Proc. Int. Symp. Electron and Positron Storage Rings*, Saclay, II – 1 – 1 (1966); *Atomnaya energiya* 22, 346 (1967).
2. Budker, G. I. et al. *Particle accelerators* 7, 197 (1976).
3. Budker, G. I. et al. *Proc. 18th Int. Conf. High Energy Physics*, Tbilisi, 86 (1976).
4. Budker, G. I. & Skrinsky, A. N. *Uspekhi Fizicheskikh Nauk*, 124, 561 (1978).

articles

Bent beams and the overall size of extragalactic radio sources

A. C. S. Readhead, M. H. Cohen & T. J. Pearson

California Institute of Technology, Pasadena, California 91125

P. N. Wilkinson

Nuffield Radio Astronomy Laboratories, Jodrell Bank, Macclesfield, Cheshire, UK

In a small sample of extragalactic radio sources, curvature is found to be a common feature in bright radio components coincident with quasi-stellar objects. In 3C273 and 3C345 the curvature is smooth between the nuclear ($<0.003''$) and the outer ($>1''$) regions. This contrasts sharply with objects having comparatively faint central radio components, which are often well aligned with the extended features.

ONE of the most significant facts to emerge from VLBI studies of extragalactic radio sources is the high degree of alignment between the most compact and the most extended features. The objects in which good alignment has been seen are 3C111, 3C236, 3C390.3, 3C405 and NGC6251 (refs 1–5). These are all 'classical double' radio sources with steep spectra and large overall size (>100 kpc). In contrast to this, it has been found^{6,7} that on different scales some objects exhibit a range of position angles. It has become increasingly clear that, in these objects, the occurrence of non-aligned structure is strongly correlated with other important properties. This effect has been noted independently by Davis *et al.*⁸

Here we describe new observations which show that in some objects the position angle changes systematically with distance from the nucleus, through an angle $\geq 20^\circ$, that is the objects are curved. The observations enable us to rule out several possible explanations of the differing position angles. We discuss two models which are consistent with all the observations and which account naturally for some of the properties of these objects. These two models have widely different implications for the physics of extragalactic radio sources.

Straight and curved jets

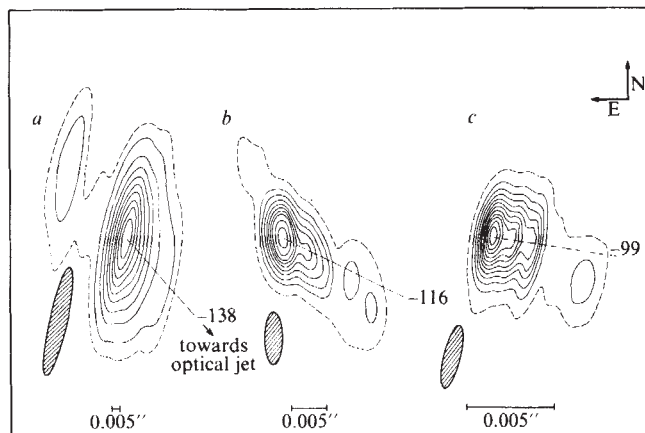
Extragalactic radio sources can be divided into two major morphological groups: the first group consists of those objects in which the dominant radio emission comes from two components situated on opposite sides of the optical object and roughly equidistant from it. There may also be a radio component coincident with the optical object, but the total emission from it is small compared with that from the outer lobes. In general, these objects are large (≥ 100 kpc overall) and, at frequencies between ~ 200 MHz and ~ 5 GHz, they have steep spectra, $-1.0 \leq \alpha \leq -0.5$, where $S \propto \nu^\alpha$. We will refer to objects in this group as 'symmetric' or 'type S' objects, to emphasise that the dominant radio emission is roughly symmetrically placed on either side of the optical nucleus. The second group consists of objects in which there is a dominant compact ($<1''$) radio

component coincident with the optical object, and if there are extended radio features these lie to one side of the optical object. These objects are generally small (≤ 100 kpc) and have flat spectra, $\alpha \approx -0.5$, in the above frequency range. We will refer to objects in this group as 'core' or 'type C' objects, to emphasise that it is the coincidence of a dominant radio component with the optical nucleus which distinguishes them from type S objects.

Recent VLBI observations at 609, 1671, 5011 and 10,651 MHz not only confirm the different position angles seen on different scales in some objects, but also show that in at least two cases, 3C273 and 3C345, there is a smooth change in position angle between the compact and extended features. The details of the observations and analysis will be given elsewhere^{9–11}.

Figure 1 shows hybrid maps of 3C273 with three different resolutions made using the technique described by Readhead and Wilkinson¹². The straight line through the brightest feature in each map, which we shall call the core, is drawn in the position angle of the feature closest to the core. Figure 1 shows that the

Fig. 1 Hybrid maps of 3C273 at three frequencies: *a*, 609; *b*, 5,011; *c*, 10,651 MHz; showing the dominant structure on three different angular scales. The ellipses show the sizes (FWHM) of the restoring beams. The dashed line in each case is drawn through the radio core and the components closest to the core. The curvature of the source is clearly seen in both (*b*) and (*c*) maps. The systematic rotation of position angle from the smallest structure (*c*) to the largest structure (609 MHz map (*a*)) is also clear.



position angle of the radio structure in 3C273 rotates systematically with distance from the core. Figure 1a and b also shows features on the opposite side of the core, but further observations are needed to confirm these. The position angle of the larger-scale (~ 30 m arc s) structure agrees, to within the errors of observation, with that of the optical jet, which extends 20 arc s away from the nucleus. The total change in position angle, between the radio core and the optical jet, is 40° , over a distance of 70 kpc, assuming $H = 50 \text{ km s}^{-1} \text{ Mpc}^{-1}$ and $q_0 = 1/2$. Thus 3C273 is a type C object which shows pronounced curvature in the radio component coincident with the optical nucleus.

Four hybrid maps of 3C345 with different resolution are shown in Fig. 2. These again show a systematic change of position angle with changing resolution. In addition, Davis *et al.*¹³ have discovered an elongated component in 3C345 which is 2 arc s away from the core in position angle (PA) -31° . This fits in well with the trend displayed by the source on smaller scales. Thus for 3C345 $\Delta\text{PA} = 45^\circ$.

In both 3C273 and 3C345 the curvature decreases as we move outwards from the core. Furthermore, the large-scale (>1 arc s) features in both cases are straight, that is they do not show curvature such as is seen near the core, and therefore, in both position and shape, these larger regions look like extensions of the compact inner regions shown in Figs 1 and 2.

These two sources provide the best evidence for a systematic rotation of position angle with distance from the radio core, but there are others which show the same tendency. For example, the 609-MHz hybrid map of 3C147 (ref. 6) shows a long narrow jet which is distinctly curved. This curvature is confirmed in a more recent map at 1,671 MHz (ref. 10). The large-scale ($\sim 0.7''$) structure¹⁴ also has a position angle different from that of the jet. The total change in position angle in 3C147 is 25° . Similarly, in 3C380 comparison of the 1,671 MHz hybrid map¹⁰ with the Cambridge 5-km telescope map at 15 GHz (ref. 15) shows that $\Delta\text{PA} = 20^\circ$.

ΔPA , overall size, distance and morphology

Our observations of compact radio sources show that changes in position angle are common in type C sources, whereas as yet no example has been found of a non-aligned type S source. The data are given in Table 1 for nine sources, in order of increasing redshift. Figure 3 shows ΔPA plotted against the overall size, D , for each source. The sharp division in both ΔPA and D between core and symmetric objects is immediately obvious. These data show that the degree of alignment of the central radio objects is closely related to the overall size, the morphology and the distance class of extragalactic radio sources. The data shown in Table 1 should be viewed with some caution, however, as the sample of sources is not complete. But with the data presently available, the correlation between these four properties is very strong.

For this discussion we will assume that there is, in all cases, a unique source of high energy particles or radiation in the optical nucleus. We assume further that the energy is transferred to the outer components by a beam^{16,17}.

We have seen that type S objects tend to be large (>100 kpc) and to exhibit good alignment between compact and extended regions, while type C objects are curved and small (<100 kpc). The present observations are crucial, when seeking an explanation of these properties, because they enable us to eliminate some of the possible causes of non-alignment in type C objects. Rotation or precession of the energy source, or the possibility of two non-aligned beams, can be ruled out by the narrowness and alignment of the outer jets (3C273 and 3C345) or by the smooth curvature in 3C273 (Fig. 1). Similarly, curvature due to acceleration of the source by a nearby galaxy, as in 3C31 (ref. 18), would not explain why the curvature in 3C273 and 3C345 is greatest nearest the core. This would imply that the maximum acceleration, and hence the closest approach of the interacting

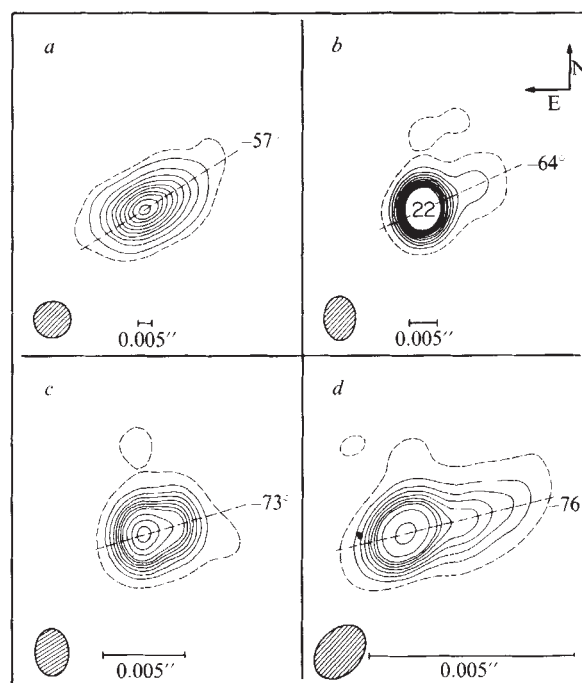


Fig. 2 Hybrid maps of 3C345 at four frequencies: a, 609; b, 1,667; c, 5,011; d, 10,651 MHz; showing the systematic rotation of position angle with structure on different scales.

galaxy, has occurred within the past few decades, which is very unlikely.

Two possible explanations for the curvature are interesting because they can also explain the smaller overall sizes of type C objects.

(1) Beam bending by pressure: if the initial direction of the beam is not along the pressure gradient of the interstellar matter the beam could be bent by external pressure forces. The physical details of this mechanism have not been worked out, and the beam may be broken up by instabilities before bending through an appreciable angle. However, if it proves possible to bend the beam without destroying it, pressure bending becomes an attractive possibility because it could also explain many of the other observed features.

The bending of the beam by the external medium creates 'working surfaces' along curved portions of the beam. Thus we expect much greater interaction between the beam and the interstellar matter than in the case of a straight beam, which has only one working surface at the end. In sources bent by pressure we therefore expect to see 'hot spots' or knots along the beam, and we expect the brightest hot spots to coincide with the most strongly curved portions of the beam. This is entirely consistent

Table 1 Morphology, overall size, change in position angle and redshift of the objects in the sample*

Object	Z	D (kpc)	Type	ΔPA (deg)
NGC6251	0.023	2,900	S	≤ 4
3C111	0.0485	270	S	≤ 4
3C390.3	0.0561	320	S	≤ 5
3C405	0.0565	190	S	≤ 4
3C236	0.0989	5,700	S	≤ 3.5
3C273	0.158	70	C	40
3C147	0.545	5	C	25
3C345	0.594	25	C	45
3C380	0.691	40	C	20

Data either taken from text or from refs 1-5.

* Assuming $H = 50 \text{ km s}^{-1} \text{ Mpc}^{-1}$ and $q_0 = 1/2$.

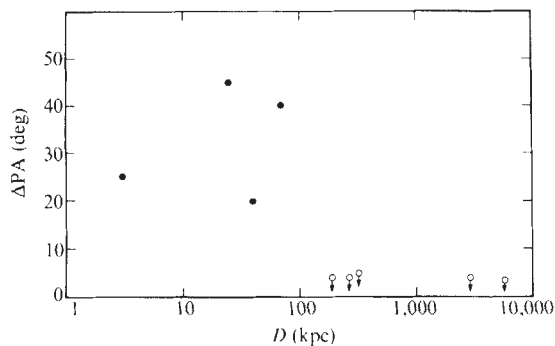


Fig. 3 The change in position angle (ΔPA) versus overall size (D) for the sources listed in Table 1: ○, type S sources; ●, type C sources.

with the observations of 3C273 and 3C345. Furthermore, much of the energy of a beam bent by external pressure forces would be dissipated inside the galaxy, and thus the beam would not travel as far. This could explain the smaller overall sizes of the curved sources and the bright synchrotron self-absorbed knots which we see in them⁹. In addition, on this model the curvature is greatest in the region of highest pressure gradient. We have seen, in 3C273 and 3C345, that the curvature is greatest at the core. This suggests the plausible explanation that the radio core is in the densest part of the galactic nucleus.

What causes the misalignment between the beam and the pressure gradient? This depends on the nature of the energy source. A strong possibility is a massive ($\geq 10^8 M_\odot$) black hole¹⁹⁻²¹, with the beam directed along the spin axis. As the black hole is formed from matter accreted from the galaxy, the angular momentum vector of the black hole, Ω_s , will be parallel to that of the nearby matter in the galaxy Ω_g . A strong gravitational interaction with another galaxy, which involves some transfer of matter, would change Ω_g to Ω'_g say, but leave Ω_s unaltered. Thus the angular momentum vector of the galaxy, and hence the pressure gradient, is no longer parallel to the beam. However, the black hole is accreting matter from the galaxy at a rate $\sim 1 M_\odot \text{ yr}^{-1}$ (ref. 21), and after $\sim 10^7$ yr, by conservation of angular momentum, Ω_s will become aligned with Ω'_g . This scheme can also explain the good alignment observed in symmetric objects. We have seen that type S objects are larger, and therefore presumably older, than type C objects. Thus it may be that the good alignment in symmetric objects is related to their age. For example, Condon and Dressel²² have shown that there is an increased probability of having a compact radio source in the nucleus of a galaxy if it is a member of a small group, or even a pair, of galaxies. Thus interaction with a nearby galaxy may trigger the production of a radio source as well as change Ω_g . This could explain why small (young) objects are curved whereas large (old, $> 10^7$ yr) objects are well aligned.

(2) Projection effects: these must be present to some degree in every object and hence affect the values of ΔPA and D which we derive from the maps. They cause us to underestimate the total object size, and can lead to either an underestimation or an overestimation of ΔPA . They can give a natural explanation of both the curvature which we observe in type C objects and the alignment of type S objects, if we assume that all of these are formed by gently curving beams in which the total deviation from a straight line is very small ($\leq 5^\circ$), and that in core objects the angle between the beam and the line of sight is small. A slightly curving beam appears strongly curved for most lines of sight almost parallel to the beam. If the radiating particles are moving at constant relativistic speeds along a narrow beam, the projection effects can also explain why the curvature is greatest nearest the core: the radiation is strongly beamed in the forward direction and the positions of maximum brightness and maximum curvature coincide as these both occur at the point where the beam is most nearly parallel to the line of sight. Therefore the majority of nuclear jets detected in a flux-limited sample will

be directed towards the observer, and any intrinsic curvature will be strongly amplified by the projection effect. The one-sided structure in 3C273 and 3C345 shows that the radio core coincides with the origin of the beam. If this were not the case the maximum curvature and brightness would occur at an intermediate position along the visible portion of the beam.

The visible features in the maps all have roughly comparable flux densities, and, because the enhancement of apparent brightness by relativistic beaming is a very strong effect, this implies that all features in the maps must be moving at relativistic speeds towards the observer. If this were not the case, the intrinsic luminosities, over 4π sr, of the stationary and moving features would have to differ by large factors which fortuitously cancel the effect of relativistic beaming. Thus the intrinsic luminosity of the central radio sources in core objects is low, and they are not as different from the central sources in symmetric objects as they appear. In symmetric objects the beam is significantly offset from the line of sight which helps to explain why the type S objects in our sample are generally nearby whereas the type C objects are more distant, because we have to observe a large volume of space to find a significant number of objects with beams pointing towards us. Furthermore, the superluminal velocities²³, like those observed in 3C273 and 3C345, are expected on this model in sources in which the angle between the line of sight and the beam is small. This is true even if all the features are moving towards the observer at relativistic speeds. The relative motion which is observed could be due to the bending of the beam, as the apparent velocity depends on the angle between the line of sight and the beam, or it could be explained if the various features have differing velocities.

Projection effects alone cannot account for all of the observed differences between core and symmetric objects. This point is considered by Scheuer and Readhead²⁴, who show that some of the remaining differences can be removed by a simple extension of the model. The details of this model and the further implications for superluminal sources and for extragalactic objects in general are discussed elsewhere^{24,25}.

We suggest that one of the above possibilities, either projection of beam bending by the external medium, is the dominant effect responsible for the curvature in most type C objects. However, it is not possible to discriminate between these alternatives with the present observations. This may become possible in the future if we can follow the progress of individual radio knots along the beam. For example, if projection effects dominate, and if the radiating particles are moving with relativistic bulk velocity, v , along the beam, the apparent transverse velocity of a knot is a maximum at an angle θ to the beam, where $\cos \theta = v/c$. Thus, in a slightly curved beam the apparent transverse velocity passes through a maximum at the point where the angle between the line of sight and the beam is closest to θ . One might therefore see the knots speeding up as they approach this point and slowing down as they move away from it. Observations of such behaviour would be strong evidence in favour of the projection argument, although it might be possible to arrange conditions in the external medium to mimic the effect in a beam bent by pressure forces. Conversely, if the velocity of the knots is found to be constant over a substantially bent portion of the beam this would be strong evidence against the projection argument.

Conclusion

We have shown that the occurrence of non-aligned radio structure in the nuclei of extragalactic radio sources is strongly correlated with three other important properties—morphology, overall size and redshift. Our observations have enabled us to rule out several possible explanations for the misalignment of type C objects. Two plausible explanations consistent with the observations are external pressure bending and projection. These are totally different in many important physical aspects. If the curvature is due to external pressure forces then the bulk velocity of the radiating particles along the jet could be non-relativistic, and the jet one-sided, that is we see all of the object.

If we are observing a projection effect in a jet with relativistic bulk motion, the jet could actually be two-sided, that is we may only see half of the object. The projection effect can only amplify a pre-existing curvature, however, and pressure may be responsible for this curvature.

Both models can explain the coincidence of maximum curvature with maximum brightness, but the projection argument is more attractive because it also accounts automatically for the superluminal velocities observed in some type C objects and the correlation of curvature with distance.

We thank M. J. Rees, W. L. W. Sargent, S. F. Gull and R. D. Blandford, for useful discussions. P.N.W. thanks the Royal Society for a Fellowship supported by the Weir Foundation. This work is supported by NSF grant AST77-00247.

Received 23 October; accepted 16 November 1978.

1. Kellermann, K. I. *Physica Scripta* **17**, 257 (1978).
2. Willis, A. G., Strom, R. G. & Wilson, A. S. *Nature* **250**, 625-630 (1974).

3. Preuss, E., Kellermann, K. I., Pauliny-Toth, I. I. K. & Shaffer, D. B. (in preparation).
4. Kellermann, K. I., Clark, B. G., Niell, A. E. & Shaffer, D. B. *Astrophys. J. Lett.* **197**, L113-L116 (1975).
5. Readhead, A. C. S., Cohen, M. H. & Blandford, R. D. *Nature* **272**, 131-134 (1978).
6. Wilkinson, P. N., Readhead, A. C. S., Purcell, G. H. & Anderson, B. *Nature* **269**, 764-768 (1977).
7. Wilkinson, P. N., Readhead, A. C. S., Anderson, B. & Purcell, G. H. *Astrophys. J.* (submitted).
8. Davis, R. J., Stannard, D. & Conway, R. G. *Mon. Not. R. astr. Soc.* (in the press).
9. Readhead, A. C. S., Pearson, T. J., Cohen, M. H., Ewing, M. S. & Moffet, A. T. *Astrophys. J.* (submitted).
10. Readhead, A. C. S. & Wilkinson, P. N. (in preparation).
11. Pearson, T. J., Readhead, A. C. S. & Wilkinson, P. N. *Astrophys. J.* (submitted).
12. Readhead, A. C. S. & Wilkinson, P. N. *Astrophys. J.* **223**, 25-36 (1978).
13. Davis, R. J., Stannard, D. & Conway, R. G. *Nature* **267**, 596-597 (1977).
14. Donaldson, W. & Smith, H. *Mon. Not. R. astr. Soc.* **151**, 253-258 (1971).
15. Scott, M. A. *Mon. Not. R. astr. Soc.* **179**, 377-388 (1977).
16. Blandford, R. D. & Rees, M. *Mon. Not. R. astr. Soc.* **169**, 395-415 (1974).
17. Scheuer, P. A. G. *Mon. Not. R. astr. Soc.* **166**, 513-528 (1974).
18. Blandford, R. D. & Icke, V. *Mon. Not. R. astr. Soc.* (in the press).
19. Lynden-Bell, D. *Nature* **223**, 690-694 (1969).
20. Frank, J. & Rees, M. H. *Mon. Not. R. astr. Soc.* **176**, 633-647 (1976).
21. Rees, M. H. *Nature* (1978) (in the press).
22. Condon, J. J. & Dressel, L. L. *Astrophys. J.* **222**, 745-751 (1978).
23. Cohen, M. H. *et al. Nature* **268**, 405-409 (1977).
24. Scheuer, P. A. G. & Readhead, A. C. S. *Nature* (submitted).
25. Cohen, M. H. *et al. Astrophys. J.* (submitted).

Olivine fractionation equations for basaltic and ultrabasic liquids

T. H. Pearce

Queen's University, Kingston, Ontario, Canada, K71 3N6

Exact equations are derived for fractionation curves of basaltic and ultrabasic liquids. The ideal fractionation equation has the form $y = ax^K$ while equilibrium crystallisation has the form $(1/y) = Km_1(1/x) + b_1 \cdot (1 - K)$, and batch fractionation, $y = Km_2 + b_2 \cdot (1 - K)$. A ratio transformation as in Pearce diagrams allows natural data to be analysed to determine if a constant distribution coefficient fits the data: K may also be determined directly. Published data for a series of liquids fractionated from a lunar picrite parent have been analysed using the equilibrium equation. The average K value is $K = 0.323$ with a linear correlation coefficient, $R = 0.996$. The reported olivine analyses yield a $K = 0.321$ for the same set of experiments. Using the ideal fractionation equation, the liquid line of descent can be calculated and constraints placed on the primitive liquids which produced a given mafic rock by olivine fractionation.

GIVEN the chemical composition of a series of rocks allegedly representing liquids which have fractionated olivine from a single parent, we should be able to use fractionation curves of the type introduced by Bowen¹, and Osborne and Schairer² to test this hypothesis. However, until the present work, the equations of exact fractionation curves requiring only chemical data had not been developed. This problem may be treated by rather complicated equations involving the fraction of material transferred from liquid to solid (see ref. 3). Of course, in dealing with natural data, the fraction of liquid converted to solid is not generally known and must be estimated. These equations, although complex, may be used to generate models on a trial and error basis. Alternatively, a step-wise computer program such as developed by Roeder⁴ can approximate the fractionation curve (however, the increments must be small because the error is cumulative).

Unless the equations are known explicitly, it is difficult to judge how well the natural data fit various fractionation models,

and extremely difficult, if not impossible, to derive a model directly from the data.

Ideally one would like to express the fractionation, either ideal (Rayleigh-type) or equilibrium, in such a simple form that it can be rearranged in a linear form with predictable slope and intercept. The direct analysis of observed distribution coefficients would then be possible and trial and error methods would be unnecessary.

Here, I will develop such equations in absolute units and show how they may be used to analyse natural data to determine not only if the distribution coefficient is constant, but also to evaluate it directly from the natural data. To use the equations, the data must be transformed using a simple transformation^{5,6} which eliminates the closed array property of data expressed as a per cent (or fraction of the whole). Many petrologists may not be familiar with the properties of such a transformation, and I will briefly explain the technique below.

Some theoretical considerations

In 1968, I proposed that certain types of ratio diagrams in which the denominator is a constant (Pearce diagrams) be used to analyse data in which a system may be considered to have progressively changed composition. The actual chemical rate of change of constituents in such a case is modelled exactly by considering absolute amounts of the constituents (grams, tons, mols and so on). In so doing, we are treating chemical variables exactly like the extensive (thermodynamic) properties of thermodynamics.

Erroneous results may be obtained by considering variations in weight per cent. Harker diagrams, for example, are particularly misleading as they plot SiO wt % on the abscissa as though it were an independent variable. SiO₂ wt % is not a single variable representing the silica content of a rock, it is rather a complex variable composed of at least 11 other variables (the other components of the rock). The closed array or, constant sum property, of weight per cent data also causes difficulties for statistical manipulations⁷.

There is another problem with per cent data (and all natural chemical data). That is, the variables cannot be separated to give

the absolute units which exactly describe the system. It is extremely difficult to apply the usual mathematical techniques to a complex variable (such as wt % SiO₂), itself composed of 12 simple variables. A ratio transformation of the data produces a mathematically simpler form (we only have two variables). The simplest possible result occurs when the denominator is a constant as in Pearce diagrams. It is a characteristic of this transformation that it allows us to obtain the same results as though the variables were separated.

This method of treating data is explained as follows, if we have a relationship between y and x , say

$$y = f(x) \quad (1)$$

then

$$\left(\frac{y}{z}\right) = g\left(\frac{x}{z}\right) \quad (2)$$

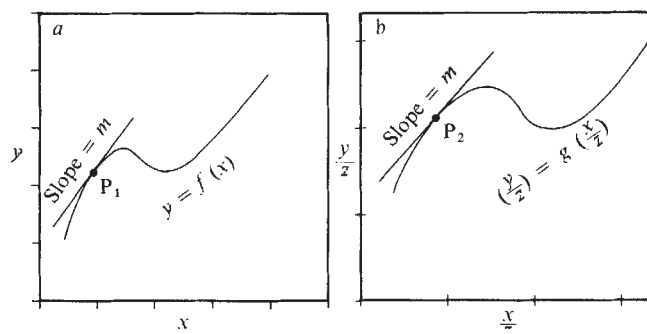
Equation (1) has the same form as (y/z) versus (x/z) , from equation (2), only if z is a constant in the mathematical sense. As $y/z = y\%/z\%$, these ratios are easily evaluated. In effect we are magnifying the curve $y = f(x)$ without changing its shape. Examining the problem graphically, we wish to obtain the relationship in Fig. 1a. We cannot do this because we cannot obtain the absolute values of y and x from closed array data (percentage values). We can, however, obtain the graph in Fig. 1b from which we have the same important information as from Fig. 1a, namely the rate of change of y with respect to x .

Every point on $y = f(x)$ has a corresponding point on $(y/z) = g(x/z)$ which bears the same relationship to y and x (that is, the slopes are the same). Thus, we can determine the true relationship between y and x by plotting $y\%/z\%$ against $x\%/z\%$ remembering that z must be a constant. This is the point of the transformation of data in a Pearce diagram and has practical applicability far beyond its mathematical simplicity. I shall use it here to calculate the distribution coefficient between olivine and liquid using only the composition of a series of liquids from which olivine has been fractionated.

Ideal fractionation equations

During ideal fractional crystallisation it is assumed that the crystals are effectively removed from the liquid as they are formed so that there is no reaction between crystals and liquid. In this case, we know, therefore, that the chemical rate of change of the liquid composition is the same as the composition of the olivine being precipitated. If we assume the distribution coefficient is constant, we can now derive the equation for ideal fractionation of olivine from the liquid as follows.

Fig. 1 a, An hypothetical graph of two related variables y and x in absolute values (grams, tons, mols). In general this type of graph cannot be determined from natural chemical data; b, as for a but with the data transformed by dividing by z , a constant. Point P_2 corresponds to point P_1 and has the same slope. In fact, as corresponding points yield the same slopes, the curve has the same shape as in (a).



Let y = amount of FeO in absolute units (such as, grams, tons, mols); x = amount of MgO in the same units as y ; K = distribution coefficient as defined by Roeder and Emslie⁸.

$$\left(\frac{\text{FeO}}{\text{MgO}}\right)_s = K \left(\frac{\text{FeO}}{\text{MgO}}\right)_L \quad (3)$$

where subscripts S, refers to solid or crystals; L, refers to liquid; i, refers to the 'i' differentiate formed; p refers to the parent. x_0 = the original amount of MgO in the system (parental MgO); y_0 = the original amount of FeO in the system (parental FeO); c_1 and so on = a constant.

Now

$$\left(\frac{y}{x}\right)_s = K \left(\frac{y}{x}\right)_L \quad (4)$$

As the crystals are physically removed from the liquid the instant they form, it follows that

$$\left(\frac{dy}{dx}\right)_L = \left(\frac{y}{x}\right)_s \quad (5)$$

considering the liquid path

$$\left(\frac{dy}{dx}\right)_L = K \left(\frac{y}{x}\right)_L \quad (6)$$

or

$$\frac{dy}{y} = K \cdot \frac{dx}{x} \quad (7)$$

integrating equation (4)

$$\int \frac{dy}{y} = K \int \frac{dx}{x} + C_1 \quad (8)$$

$$\log y = K \log x + C_1 \quad (9)$$

evaluating the integration constant C_1

$$\log y = K \log x + \log (y_0/x_0^K) \quad (10)$$

or

$$y = \left(\frac{y_0}{x_0^K}\right) \cdot x^K \quad (11)$$

These simple equations describe the exact shape of the ideal fractionation curve in absolute amounts and may be used to model olivine fractionation. They cannot be used to analyse natural data because natural chemical data cannot be separated into discrete, simple, variables. However, by using the transformation of a Pearce diagram, we can achieve the same result. Therefore, using theorem 1 of Pearce⁵: y/z versus x/z will give us the same relationship as y versus x , and therefore

$$\log \left(\frac{y}{z}\right) = K \log \left(\frac{x}{z}\right) + \log \left(\frac{y_0}{x_0^K} \cdot z^{K-1}\right) \quad (12)$$

and

$$\left(\frac{y}{z}\right) = \left(\frac{y_0}{x_0^K}\right) \cdot \left(\frac{z}{x_0}\right)^{K-1} \cdot \left(\frac{x}{z}\right)^K \quad (13)$$

Both equations (12) and (13) may be used to analyse natural data but equation (12) is more useful as it gives a straight line with slope K on a $\log(y/z)$ versus $\log(x/z)$ diagram. Thus the presence of a constant distribution coefficient as well as its value may be determined from natural liquid compositions. Rewriting equation (12) in terms of oxide constituents

$$\log \left(\frac{\text{FeO}}{\text{Al}_2\text{O}_3}\right) = K \log \left(\frac{\text{MgO}}{\text{Al}_2\text{O}_3}\right) + \log \left(\frac{\text{FeO}}{\text{MgO}}\right)_p \cdot \left(\frac{\text{Al}_2\text{O}_3}{\text{MgO}}\right)_p^{K-1} \quad (14)$$

Note that in equations (12) and (13) the exact size of the system need not be specified as the equation does not change with the size of the system.

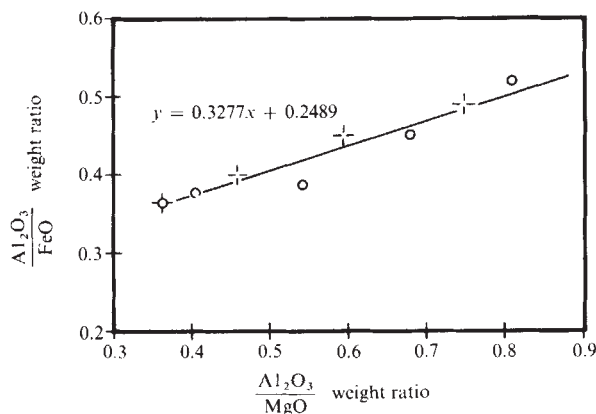


Fig. 2 Weight ratio data for fractionated liquids from a lunar picrite. The points are both actual experimental data (○) and smoothed data (+). The linear regressive line has a linear correlation coefficient of $R = 0.996$. The average K determined from both slope and intercept is 0.323. Walker *et al.* report data giving an average K of 0.321. Note that this type of 'inverted data' graph should only be used to analyse the K values for liquids which have formed by equilibrium crystallisation.

Equilibrium crystallisation equations

During equilibrium crystallisation there is continuous, complete reaction between the crystals and coexisting liquid. At any instant there is only one solid composition which represents the total amount of crystals which have formed, and the system consists of only two phases, crystals and liquid. Unlike ideal (non-reactive) fractionation, we do not know, at first sight, the chemical rate of change of the liquid (dy/dx), however, we do know that, because of the constant composition constraint, (in absolute amounts):

$$y_0 = y_s + y_L \quad (15)$$

and

$$x_0 = x_s + x_L \quad (16)$$

also

$$\left(\frac{y}{x}\right)_s = K \left(\frac{y}{x}\right)_L \quad (4)$$

therefore, combining equations (15), (16) and (4), in terms of the liquid curve

$$\left(\frac{y}{x}\right)_L = \frac{1}{K} \left(\frac{y_0 - y_L}{x_0 - x_L}\right) \quad (17)$$

As we are concerned only with the liquid, we can drop the subscripts and rearrange equation (17) by cross multiplication to give the simple form

$$\left(\frac{1}{y}\right) = K \left(\frac{x_0}{y_0}\right) \left(\frac{1}{x}\right) + \frac{(1-K)}{y_0} \quad (18)$$

This is the equation of a straight line with slope $K(x_0/y_0)$ and y -intercept $(1-K)/y_0$ if $(1/y)$ is plotted against $(1/x)$. This equation may be used for modelling but it cannot be used directly to analyse natural data because it is written in absolute amounts. Using the transformation advocated here, substituting y/z and x/z for y and x

$$\left(\frac{z}{y}\right) = K \left(\frac{x_0}{y_0}\right) \left(\frac{z}{x}\right) + \left(\frac{z_0}{y_0}\right) \cdot (1-K) \quad (19)$$

Rewriting in terms of oxide constituents

$$\left(\frac{\text{Al}_2\text{O}_3}{\text{FeO}}\right)_i = K \left(\frac{\text{MgO}}{\text{FeO}}\right)_p \cdot \left(\frac{\text{Al}_2\text{O}_3}{\text{MgO}}\right)_i + \left(\frac{\text{Al}_2\text{O}_3}{\text{FeO}}\right)_p \cdot (1-K) \quad (20)$$

These ratios, which are readily determined from chemical analyses, may be in any units and K calculated from the slope

will be the same. Of course, any element, oxide or extensive variable of the system could be used as the denominator of the ratios as long as it is constant. However, it is convenient to use Al_2O_3 or $(\text{Al}_2\text{O}_3 + \text{CaO})$, although TiO_2 is also useful (especially if one attempts to model plagioclase fractionation in addition to olivine).

Batch fractionation equations

In batch fractionation, which is essentially a disequilibrium non-reaction process, we assume a constant composition of olivine determined by the original liquid composition and distribution coefficient. The equation for such a process is

$$y = K \left(\frac{y_0}{x_0}\right) x + y_0 \cdot (1-K) \quad (21)$$

or applying the transformation as in Pearce⁵

$$\left(\frac{y}{z}\right) = K \left(\frac{y_0}{x_0}\right) \left(\frac{x}{z}\right) + \left(\frac{y_0}{z_0}\right) \cdot (1-K) \quad (22)$$

Rewriting in oxide form using Al_2O_3 as a divisor we have

$$\left(\frac{\text{FeO}}{\text{Al}_2\text{O}_3}\right)_i = K \left(\frac{\text{FeO}}{\text{MgO}}\right)_p \cdot \left(\frac{\text{MgO}}{\text{Al}_2\text{O}_3}\right)_i + \left(\frac{\text{FeO}}{\text{Al}_2\text{O}_3}\right)_p \cdot (1-K) \quad (23)$$

Analysis of liquid composition

If we have the compositions of a series of liquids produced by ideal fractionation of olivine, then from equation (14) it is clear that a plot of $(\text{FeO}/\text{Al}_2\text{O}_3)$ versus $(\text{MgO}/\text{Al}_2\text{O}_3)$ will be linear with a slope of $m = K$ and a y -intercept of

$$b = \log \left(\frac{\text{FeO}}{\text{MgO}}\right)_p \cdot \left(\frac{\text{Al}_2\text{O}_3}{\text{MgO}}\right)_p^{K-1} \quad (24)$$

Thus we have two ways of evaluating K from a single graph.

Unfortunately I do not have access to natural data to demonstrate the effectiveness of this technique. However, data are available for equilibrium crystallisation and that will be analysed below.

Table 1 Lunar picrite observed and calculated liquid compositions

	1	2	3	4
FLIQ	180	184		1250
Olivine	1.00	0.83	0.83	0.84
mol %				
Olivine	77.3	69.4	68.7	72.1
observed				
SiO ₂	—	69.7	—	—
TiO ₂	43.8	44.0	45.3	44.3
Al ₂ O ₃	2.80	3.33	3.4	3.3
FeO	7.95	9.63	9.68	9.5
MnO	22.0	21.3	21.7	21.1
MgO	0.29	—	0.35	—
CaO	13.5	8.69	8.55	9.8
Na ₂ O	8.18	9.95	9.96	9.5
K ₂ O	0.22	—	0.27	0.30
Cr ₂ O ₃	—	—	0	0.07
	0.65	—	0.79	1.0
	99.39	96.90	100.00	98.87

Analyses 1, 2 and 4 from Walker *et al.*⁹. FLIQ is the fraction of liquid remaining after fractionation. Olivine mol % is the equilibrium olivine calculated from equation (3) assuming a K of 0.32. Olivine observed is the actual olivine analysed by Walker *et al.*⁹. 1. Starting composition, lunar picrite 12002. 2. Liquid fractionated from 12002 by olivine fractionation. 83% of the liquid remains. Compare this composition with nos 3 and 4. 3. Fractionated liquid calculated using equation (18) and summing the absolute amounts to 100%. Note the close agreement with the experimental run no. 2. 4. Walker *et al.*⁹ reported this fractionated composition for 1,250 °C and 84% liquid remaining. Compare with nos 2 and 3. Note that the MgO content is slightly high resulting in a calculated olivine slightly enriched in forsterite.

In the case of equilibrium crystallisation it is clear from equations (19) and (20) that a plot of $(\text{Al}_2\text{O}_3/\text{FeO})$ against $(\text{Al}_2\text{O}_3)/(\text{MgO}) \cdot (\text{MgO}/\text{FeO})_p$ will be linear with a slope of $m = K$, and a y-intercept of

$$b = \left(\frac{\text{Al}_2\text{O}_3}{\text{FeO}} \right)_p \cdot (1 - K) \quad (25)$$

Thus, in this case also, we have two different ways of evaluating K from the same graph.

Walker *et al.*⁹ have reported experimental data for equilibrium crystallisation of olivine from a lunar picrite. These data are plotted in Fig. 3 and may be used to evaluate the nature and value of the distribution coefficient. Note that in this case the x -values are multiplied by $(\text{MgO}/\text{FeO})_p$ to yield a slope of K . Using Walker's fractionated liquid values the observed K calculated from the liquids is 0.328 (from the slope) and 0.319 (from the intercept) giving an average value of 0.323. This compares favourably with the average K of 0.321 calculated from the reported liquid and olivine compositions. Table 1, shows compositions calculated using equation (18) and summing the absolute amounts to 100%: the excellent agreement requires no further comment (see Fig. 3).

Table 2 Calculated parental liquids for an abyssal tholeiite

	1	2	3	4	5	6
FLIQ	1.000	1.183	1.355	1.521	1.685	1.846
Olivine mol%						
SiO ₂	82.2	88.2	90.9	92.42	93.4	94.2
TiO ₂	50.00	48.51	47.57	46.90	46.39	45.99
Al ₂ O ₃	1.64	1.39	1.21	1.08	0.97	0.89
Fe ₂ O ₃	15.00	12.68	11.07	9.86	8.91	8.13
FeO	0.40	0.34	0.30	0.26	0.24	0.22
MnO	10.80	11.24	11.09	10.76	10.39	10.02
MgO	0.22	0.19	0.16	0.15	0.13	0.12
CaO	8.40	14.20	18.61	22.09	24.93	27.31
Na ₂ O	10.60	8.96	7.83	6.97	6.29	5.74
K ₂ O	2.70	2.28	1.99	1.78	1.60	1.46
P ₂ O ₅	0.13	0.11	0.10	0.09	0.08	0.07
H ₂ O	0.11	0.09	0.08	0.07	0.06	0.06
	0.00	0.00	0.00	0.00	0.00	0.00

1, Abyssal tholeiite studied by Fujii and Kushiro¹⁰. The composition was adjusted to give $\text{Fe}_2\text{O}_3 = 0.40$ and summed to 100%. 2–6, Possible parental liquids for no. 1 calculated using equation (11) and summing the absolute amounts to 100%. The assumptions are: $K = 0.30$ and ideal fractionation of olivine. Composition no. 4 has 52.1% olivine added and is in equilibrium with olivine of $\text{Fo}_{92.4}$. Composition nos 1–6 define the olivine-controlled liquid line of descent which leads to the abyssal tholeiite no. 1.

Calculation of primitive liquids

There is another way in which the explicit equations developed here may be used in petrogenetic theory. The family of curves described by equation (11) are, in effect, olivine-controlled liquid lines of descent. Considering the form of the equations, it is clear that a single point on the curve defines the equation. Therefore, only one ideal fractionation curve passes through any given composition. The consequence of this is important. If a single differentiate is known, then all possible parental liquids may be calculated exactly. As ideal fractionation is a limiting case, limits may be placed on the compositions of peridotitic liquids which may be parental to a given basic or ultrabasic liquid, always assuming, of course, that K is constant, and known. Table 2 shows a set of primitive liquids calculated for a recently studied abyssal tholeiite¹⁰.

Unlike ideal fractionation, the equation for the family of equilibrium curves contains two parameters (in addition to K). Thus two points on any curve are necessary to define that curve. As a consequence of this, a given mafic liquid never lies on a single equilibrium crystallisation line of descent. An infinite

number of equilibrium curves pass through any given composition. Thus the constraints on possible parental compositions are more complicated than in the case of ideal fractionation. A discussion of the details of determining the parent line of composition is beyond the scope of this work, and will be reported later.

Discussion

While developing these equations, some problems of applicability to natural data became apparent. First, most petrologists are still publishing whole rock analyses of porphyritic rocks. Bowen¹¹ has pointed out that whole rock compositions do not necessarily represent liquids, hence whole rock compositions have limited value in petrogenetic studies. The modern approach would seem to be to analyse the matrix separately. Such data would give the liquid line of descent directly and be more useful in theoretical studies. Attempts to correct whole rock analyses to a phenocryst-free composition by using modal analyses are unsatisfactory. The errors introduced by the modal analyses are such as to render the resulting data virtually worthless. Second, the value of K determined from the above equations is extremely susceptible to errors in analyses. Therefore, it is suggested that in some studies replication of analyses may be useful. Alternately, smoothing the data to a curve, as in ref. 9, seems to work well, especially if there are only a few points on the line.

Although calculation of liquid composition is the main concern of the present work, calculation of the total solid composition is readily accomplished by subtracting the liquid composition at any point from the starting composition.

In spite of the different form of the equations for ideal and equilibrium fractionation, the results for differentiation of basalt are rather similar. In fact, for small degrees of fractionation they may be difficult to distinguish. However, the two equations are fundamentally different. If, for example, at any point during equilibrium crystallisation the liquid is isolated from the solid, then the system initiates a new curve from that point. For ideal fractionation, on the other hand, only a single curve exists.

It may be worth noting that the fractionation curves are independent of temperature as long as K is independent of temperature. The equations reported here, although developed for olivine fractionation, are completely general, and were developed in terms of y and x to emphasise this generality. The equations, therefore, apply not just to olivine, but to any mineral for which K is constant and which has a solid solution between two end-members. The equations apply directly to orthopyroxene and clinopyroxene (if conditions are such that K is constant). Plagioclase is a more difficult case, but similar equations are being developed for plagioclase fractionation and these will be published later.

I thank P. L. Roeder and T. N. Irvine whose work motivated me to tackle this problem, and P. L. Roeder for the results of his olivine fractionation program which greatly helped in the development of the equations. The mathematical derivations were reviewed by H. Kummer but any errors are mine. L. Dykes did the drafting and key-punching for the computer runs. This work was supported by NRC of Canada operating grant A-8709.

Received 18 July; accepted 31 October 1978.

- Bowen, N. L. *Proc. natn. Acad. Sci. U.S.A.* **27**, 301–309 (1941).
- Osborn, E. J. & Schairer, J. F. *Am. J. Sci.* **239**, 715–763 (1941).
- Irvine, I. N. *Yb. Carnegie Instn Wash.* **76**, 454–461 (1977).
- Roeder, P. L. *Fortschr. Miner.* **52**, 61–73 (1975).
- Pearce, T. H. *Contr. Miner. Petrol.* **19**, 142–157 (1968).
- Pearce, T. H. *J. Petrol.* **11**, 15–32 (1970).
- Chayes, F. *J. Geol.* **70**, 440–452 (1962).
- Roeder, P. L. & Emslie, R. F. *Contr. Mineral. Petrol.* **29**, 275–289 (1970).
- Walker, D., Kirkpatrick, R. J., Longlin, J. & Hays, J. F. *Geol. Soc. Am. Bull.* **87**, 646–666 (1976).
- Fujii, F. & Kushiro, J. *Yb. Carnegie Instn Wash.* **76**, 461–465 (1977).
- Bowen, N. L. *The Evolution of the Igneous Rocks* (Princeton University Press, 1928) (reprinted by Dover Publications, New York, 1956).

Degenerate perturbations of protein structure as the mechanism of anaesthetic action

C. D. Richards

Division of Neurophysiology and Neuropharmacology, National Institute for Medical Research, Mill Hill, London NW7, UK

K. Martin & S. Gregory

Department of Pharmacology, Medical School, Hills Road, Cambridge, UK

C. A. Keightley, T. R. Hesketh, G. A. Smith, G. B. Warren & J. C. Metcalfe

Department of Biochemistry, Tennis Court Road, Cambridge, UK

*The interaction of the *n*-alkanols with lipid bilayers and excitable membranes shows that there is no simple correlation between conduction block and any of the perturbations of bilayer structure currently proposed as unitary mechanisms of local anaesthetic action. We propose instead that the *n*-alkanols act by direct interaction with target proteins to cause perturbations which depend directly on the precise structure of the alcohol.*

A CLASSICAL EXAMPLE of nonspecific drug action lies in the effects of local anaesthetics on isolated nerve preparations. The extraordinary range of chemically diverse structures which will cause reversible conduction block suggests that there can be no specific chemical interaction between a unique receptor site on the proteins responsible for nerve conduction and all the molecules which have local anaesthetic activity. Nevertheless, it has been widely assumed as a working hypothesis that there may be a common structural perturbation in the excitable membrane which is caused by all anaesthetics irrespective of their structure and which is sensed by the target proteins¹. The underlying rationale for such unitary hypotheses derives from the experimental evidence that anaesthetic potency correlates approximately with the oil-water partition coefficient of the anaesthetics², implying that anaesthetic action occurs when about the same number of molecules interact with the excitable membrane. The assumption is that, as the bound anaesthetics at the same membrane concentration all perturb function to about the same extent, they may all do so by the same molecular mechanism³.

However, there has been no direct experimental evidence for any unitary mechanism because it has not yet been possible to analyse directly the interaction of anaesthetics with the target protein structures in the excitable membrane. For this reason many other membranes^{4,5} have been used as models to analyse the perturbation of membrane structure and function by anaesthetic molecules. Unfortunately, as a direct consequence of the nonspecific nature of the interactions, virtually every physical and biochemical parameter which can be measured may be affected, many at, or just above, the concentrations required for local anaesthetic action⁶. Despite the problem of identifying the relative importance of these many perturbations, studies have focused attention on two contrasting classes of structural perturbation as candidates for a unitary mechanism of anaesthetic action. These are perturbations of the structure of the lipid bilayer leading indirectly to perturbations of the target protein structures, or alternatively, perturbations of the target protein structures through their direct interaction with the anaesthetics.

Here we compare the effects of *n*-alkanols on various properties of the lipid bilayer with their ability to block nerve conduc-

tion, which provides a critical test to distinguish between these two main hypotheses.

Perturbations of fluidity and phase transitions by *n*-alkanols

The effects of the *n*-alkanols on the fluidity of a range of defined lipid bilayers was estimated using the spin label technique. The ratio of the low-field to mid-field peak heights (h^{-1}/h^0) was used as an arbitrary measure of bilayer fluidity to demonstrate the effects of fluidity on the short- and long-chain *n*-alkanols. As an example, Fig. 1 (see legend for details) shows the effects of butanol and dodecanol, separately and in combination, on dioleoyl phosphatidylcholine (DOPC) and oleoyl palmitoyl phosphatidylcholine (OPPC) with and without cholesterol, for spin-labelled phospholipids carrying the nitroxide group at C-16. These lipid mixtures were selected to span the range of bilayer fluidity likely to be found in excitable membranes. Similar results were obtained when the spin label was in the C-7 or C-10 position (not shown).

In general, a similar pattern of opposing perturbations and antagonistic effects of the two *n*-alkanols in combination is obtained irrespective of the composition of the lipid bilayer or the spin label used. In detail, the magnitude of the effects varies, in that the decrease in fluidity caused by dodecanol is small in bilayers containing high proportions of cholesterol, which is itself a potent rigidifying agent, and the spectral changes observed with the nitroxide at the C-7 position are marginal with either alkanol. Similar results are obtained with other pairs of short- and long-chain *n*-alkanols; *n*-alkanols with a chain length of <10 carbon atoms increase membrane fluidity whereas those with a chain length >10 carbon atoms decrease it (not shown). The general conclusion, that the short- and long-chain *n*-alkanols cause opposing changes in bilayer fluidity holds not only for the defined lipid mixtures, but also for bilayers of lipids extracted from erythrocytes (not shown).

We assume that the same conclusion will probably apply to the effects of the alcohols on the fluidity of the lipid bilayer in excitable membranes, particularly as the lipid composition of the axonal membranes closely resembles that of the erythrocyte membrane^{7,8}. A similar pattern of responses emerges from the effects of the *n*-alkanols on the transition temperature ($T_i = 23.5^\circ\text{C}$) of dimyristoyl phosphatidylcholine (DMPC) (Fig. 2a); the short-chain alkanols depress T_i , whereas the long-chain alkanols increase T_i , as reported previously^{9,10}. The effect of butanol on DMPC-dodecanol mixtures (2:1 and 1:1) shows the consistent antagonistic action of the combined alkanols on T_i (Fig. 2b). Similar results were obtained with dipalmitoyl phosphatidylcholine, and the increase in T_i with the long-chain alkanols does not therefore depend on any precise relationship between the chain length of the alkanol and that of the lipid.

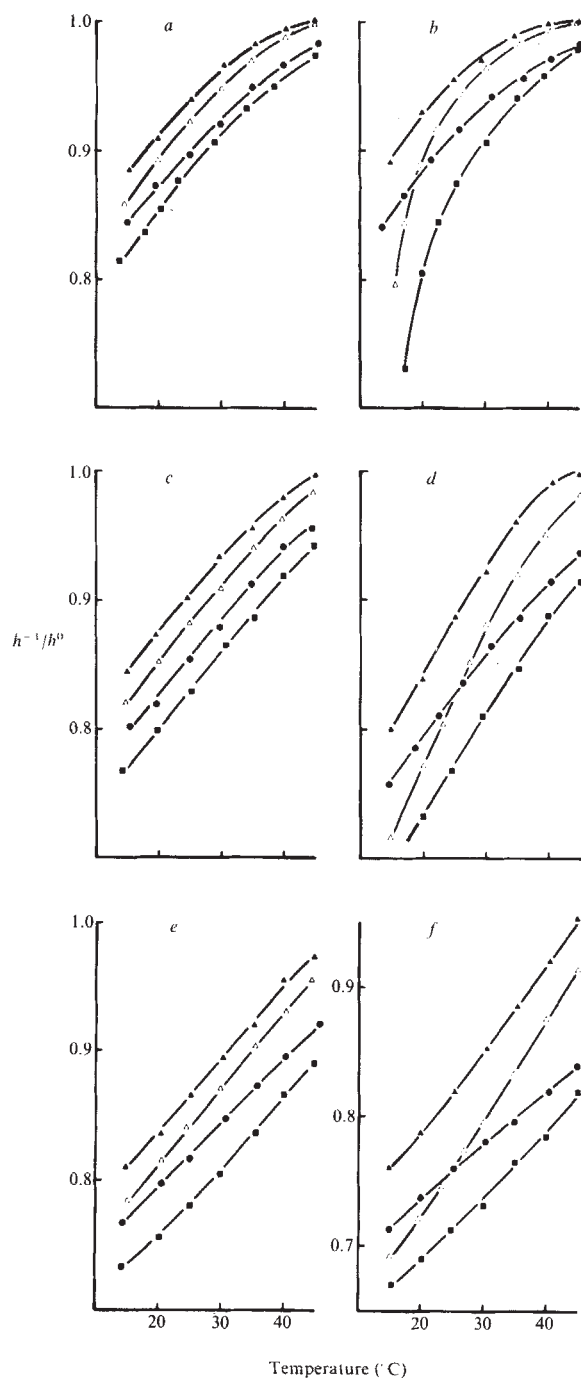


Fig. 1 The effects of butanol (C₄) and dodecanol (C₁₂) on the fluidity of dioleoyl phosphatidylcholine (DOPC) and palmitoyl phosphatidylcholine (OPPC) bilayers, and mixtures of these lipids with cholesterol (Ch). The phospholipids, synthesised by the method of Robles and van den Berg³⁶, were dissolved in CHCl₃ with spin label (1 mol%), with or without added cholesterol and C₁₂. The samples were dried under N₂ and resuspended in 100 mM potassium phosphate buffer, pH 7.2, to give a final concentration of 75 mg ml⁻¹ phospholipid. For experiments in 150 mM C₄, the alkanol was added to the buffer before resuspension of the dried lipid samples. The spin labels used were the *N*-oxyl-4',4'-dimethyl-oxazolidine derivatives of 1-palmitoyl-2(*n*-ketostearoyl) phosphatidylcholine (*n* = 7, 10 or 16) (results for *n* = 7, 10 not shown). Spectra were recorded on a Varian E3 EPR spectrometer with a variable temperature controller. The fluidity was estimated arbitrarily from the ratio of the low-field to mid-field peak amplitudes (h^{-1}/h^0) (ref. 37), where an increase in this ratio indicates an increase in bilayer fluidity. *a*, DOPC; *b*, OPPC; *c*, DOPC/Ch, 4/1; *d*, OPPC/Ch, 4/1; *e*, DOPC/Ch, 2/1; *f*, OPPC/Ch, 2/1. ▲, C₄ (150 mM); △, C₄ + C₁₂; ●, no alkanol; ■, C₁₂ (2 lipid : 1 alkanol).

Nerve-blocking action of alkanols

The local anaesthetic actions of the primary *n*-alkanols were investigated using an *in vitro* preparation of the lateral olfactory tract (LOT) of the guinea pig. This nerve consists of thinly myelinated nerve fibres of uniform size (1–2 μm diameter) which conduct impulses at 6–8 ms⁻¹ (refs 11, 12). We chose this for our experiments as it does not possess the layers of connective tissue characteristic of peripheral nerves that may present a barrier to the diffusion of the less soluble alkanols.

A well defined biphasic compound action potential of 1–6 mV amplitude and 1.5 ms duration was recorded along the length of the nerve after a supramaximal stimulus to the anterior end of the LOT¹³. On exposure to butanol or octanol the first indication of the start of nerve block was an increase in both the latency and duration of the compound action potential. Thereafter, the amplitude of the action potential declined progressively and the latency and duration increased still further, until the nerve was completely blocked. For concentrations of butanol and octanol just sufficient to block nerve conduction this sequence of events took about 15 min (Fig. 3). The actions of the long-chain *n*-alkanols (C₁₀–C₁₂) were similar to those of butanol and octanol. The time taken to achieve nerve block increased progressively as the water solubility of the *n*-alkanols diminished, being up to 700 min for dodecanol (Fig. 3). In the absence of anaesthetic the fibres of the LOT conducted normally for periods in excess of 12 h. Conduction block by all the *n*-alkanols up to C₁₁ was shown to be reversible. The effect of dodecanol could only be partially reversed even when nerve block was incomplete (Fig. 3*b*). The perfusion times required to remove the *n*-alkanols increased progressively with increasing chain length and decreasing water solubility.

It has been shown that short-chain alkanols (C₂–C₈) block nerve impulse conduction in the squid giant axon by inhibiting the increase in sodium conductance that is associated with the nerve impulse¹⁴. This also seems the most likely explanation for the action of the alkanols on the LOT, especially as their effects were similar to the action of tetrodotoxin, a specific sodium-channel blocker. However, nerve block would also be predicted if the alcohols acted by an inhibition of the sodium pump¹⁵. To test this, nerves were exposed to the sodium-pump inhibitor ouabain (10–15 μM)¹⁶. Low concentrations of ouabain (10 μM) produced a reversible block of the action potential within 20 min. The rate of decline of the action potentials could be accelerated by repetitive stimulation, as one would expect if this effect were due to block of the sodium pump. In contrast, the effects of the alkanols could not be enhanced by repetitive stimulation (not shown). It therefore seems that both short- and long-chain alkanols act directly on the sodium conductance mechanism and not on the sodium pump.

We have already noted that the long-chain and short-chain *n*-alkanols have opposing effects on membrane fluidity (Figs 1, 2), but both block nerve conduction. Thus, there is no direct correlation between the effects of the *n*-alkanols on membrane fluidity or on phase transitions and their ability to block the conduction of nerve impulses. It could, however, be argued that the decrease in fluidity caused by long-chain alkanols (C₁₀–C₁₂) is responsible for their nerve-blocking action⁶, and if this were so the anaesthetic action of all the alkanols could still be regarded as the result of perturbations of the membrane lipids. This hypothesis predicts that alkanols which have opposing actions on membrane fluidity and phase transitions should also have antagonistic actions in their local anaesthetic properties⁶.

This proposition was tested by studying the interaction between the local anaesthetic actions of butanol and dodecanol in the LOT. Conventional methods of isobolic analysis of drug interaction were not practicable because of the slow time course of the action of dodecanol. Instead, the depressant action of 30 mM butanol on the LOT compound action potential was measured at various times during continuous perfusion of the nerve with a weak suspension of dodecanol (0.2 mM) in artificial cerebrospinal fluid (CSF). If the actions of the two alkanols were

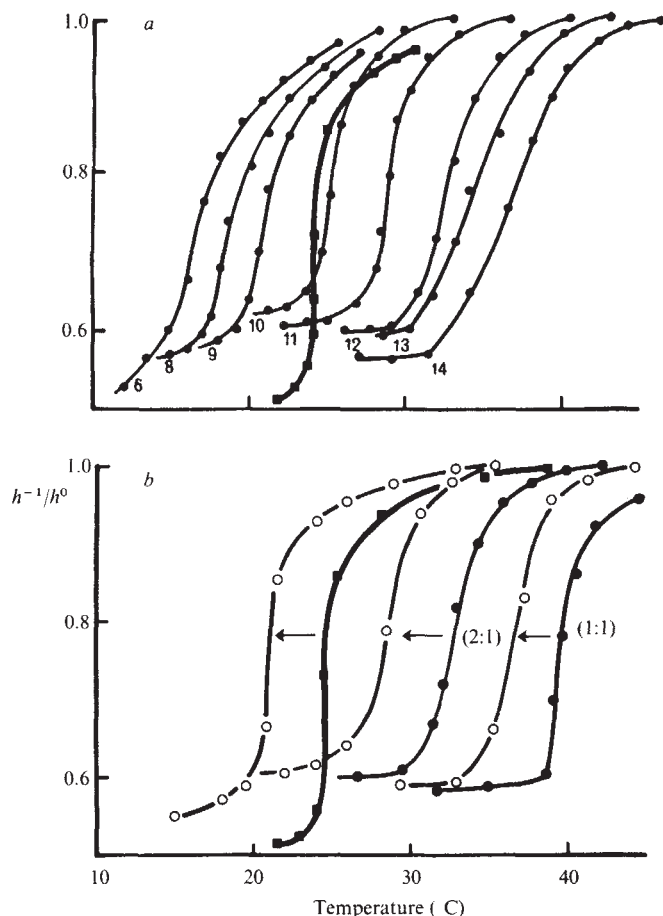


Fig. 2 *a*, The effects of *n*-alkanols on the phase transition of dimyristoyl phosphatidylcholine (DMPC). The phase transition of an aqueous suspension of pure DMPC is shown by the thick line and solid square symbols. All the other curves refer to the phase transition in the presence of alcohols; the chain length is given at the foot of each curve. The alkanols were added either as an aqueous suspension in phosphate buffer to the dried-down lipid samples with the spin label 4',4'-dimethylaxazolidine derivative of 16-ketostearic acid as in the legend to Fig. 1, or were mixed directly with the lipid in chloroform before drying down. The final alkanol concentrations were 80 mM to C_9 or 50 mM for C_{10} – C_{14} , and the DMPC concentration was 100 mM. *b*, The effect of butanol (C_4 , ○) and dodecanol (C_{12} , ●) on the phase transition of DMPC separately and in combination (DMPC: C_{12} at 1:1 or 2:1 molar ratios). As in *a* the DMPC phase transition curve (■) is depressed by butanol and raised by dodecanol. Butanol antagonism (arrows) of the elevation of the phase transition by dodecanol is shown for 1:1 and 1:2 ratios of dodecanol to lipid.

antagonistic, the depressant effect of the butanol would be expected to decrease as the nerve took up the dodecanol. However, the effectiveness of butanol increased progressively with the period of exposure of the nerve to dodecanol (Fig. 4) and this increase in potency was apparent long before the depressant action of dodecanol itself became evident. A similar effect, of even slower time course, was seen when a nerve was exposed to tridecanol instead of dodecanol. The increase in the potency of butanol was not attributable to deterioration of the nerve with time, because the effect of repeated applications of 30 mM butanol to nerves superfused only by artificial CSF was essentially constant over 12 h (Fig. 4).

Other chemical agents have been found to antagonise the physical effects of the short-chain alkanols on lipid bilayers and membranes. For example, the alkanols protect red cells from hypotonic haemolysis, whereas the pyridine derivative 4,*N*-ethylaminoethyl,2-methyl,3-hydroxy,5*S*-

methyl-thiomethyl pyridine (EMD-21657; Merck Darmstadt) antagonises the anti-haemolytic actions of alkanols and is an effective treatment of alcoholic organic brain syndrome⁴¹. Although EMD-21657 alone had no effect on the transition temperature of dimyristoyl phosphatidylcholine at up to equimolar concentrations with the lipid, it slightly antagonised the effect of hexanol in lowering the transition temperature of the lipid. However, EMD-21657 consistently potentiated the local anaesthetic action of hexanol on the LOT (Fig. 5) and, by itself,

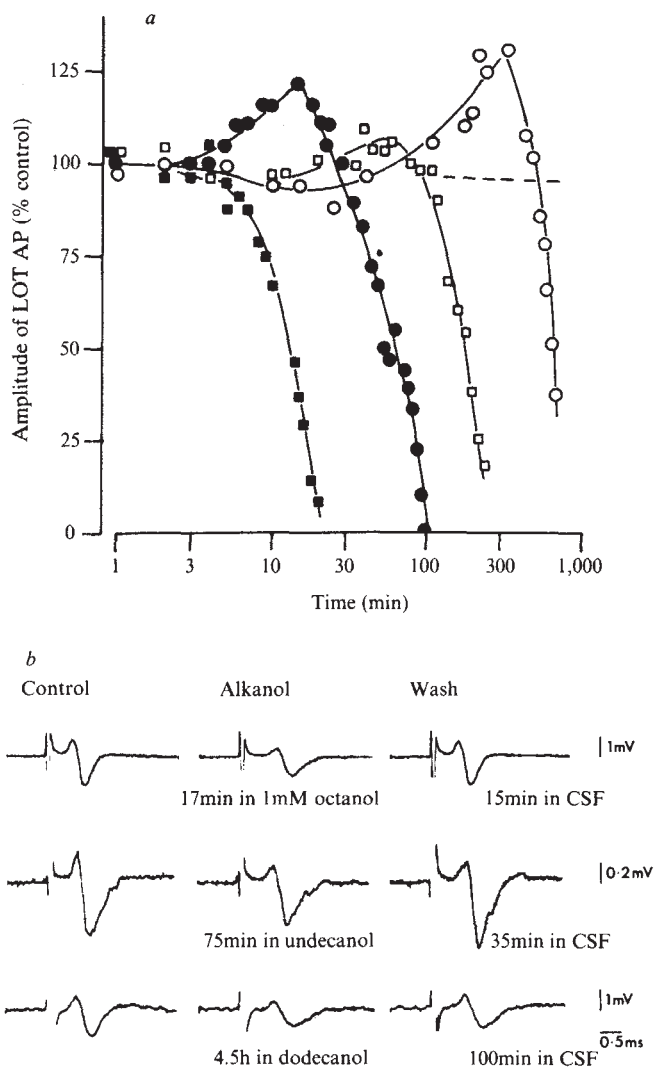


Fig. 3 Local anaesthetic action of primary alkanols. *a*, Time course of action; each alkanol was administered continuously from zero time as a solution or suspension. The ordinate represents the amplitude of the LOT compound action potential expressed as a percentage of its amplitude immediately before the administration of the alkanol. Note logarithmic time scale along abscissa. Results from four separate experiments are shown. Those of a fifth nerve exposed to CSF only are indicated by the broken line (the experimental points were omitted for clarity as were the results for butanol which were essentially similar to those for octanol). ■, C_8 ; ●, C_{10} ; □, C_{11} ; ○, C_{12} ; ---, CSF. *b*, Sample records from three experiments. As the alkanols affect nerve conduction the action potential becomes smaller and broader, and occurs at a longer latency. This combination of effects is similar for all alkanols and other anaesthetics. Note that the reversal of nerve block by washing with CSF becomes slower with increasing chain length. The details of preparation and incubation have been described previously^{38,39}. Octanol was administered as a 1 mM solution in artificial CSF and the long-chain alcohols were administered as suspensions of 1 mM for decanol, 0.5 mM for undecanol and 0.2 mM for dodecanol in CSF. All experiments were carried out at 38 °C with supramaximal stimuli given to the LOT at 0.2 Hz.

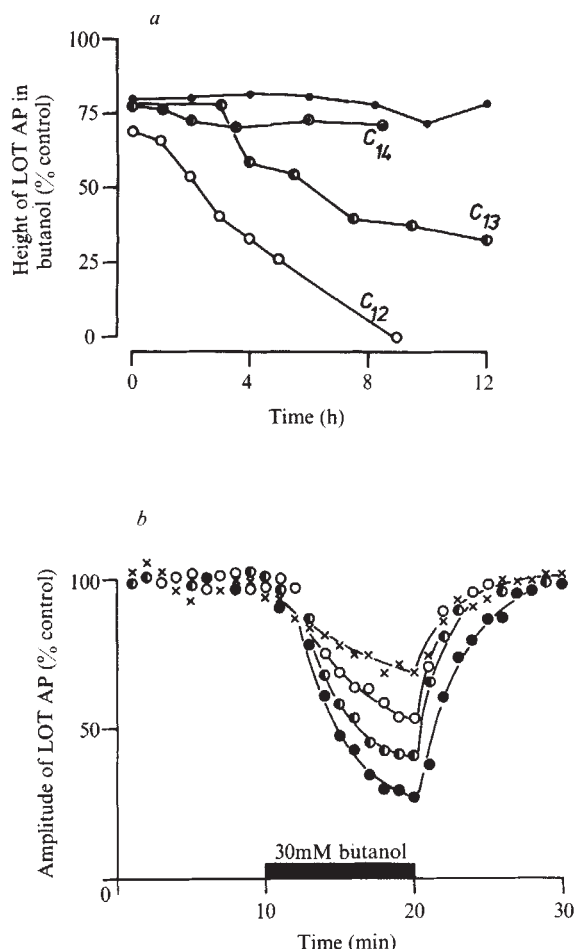


Fig. 4 Interaction between the local anaesthetic actions of butanol and long-chain alkanols (C₁₂–C₁₄). Experimental details as for Fig. 3. *a*, Four representative experiments are shown. Each nerve was exposed to 30 mM butanol for 10 min at various times while a continuous application of one long-chain alkanol was given as indicated. The ordinate represents the amplitude of the LOT action potential (AP) after 10 min exposure to 30 mM butanol expressed as a percentage of the average amplitude for the 5 min before exposure. Dodecanol (0.2 mM) and tridecanol (0.1 mM) were administered as suspensions, tetradecanol was administered at 50 μ M in a liposomal suspension (molar ratio of 1:2 with egg yolk lecithin prepared as previously described²⁸). ●, No addition; ○, C₁₄; ○, C₁₃; ○, C₁₂. *b*, An example of the potentiating action of dodecanol on the local anaesthetic action of butanol. Initially, the nerve was exposed to 30 mM butanol to determine its sensitivity (x). After washing out the butanol the dodecanol suspension was applied continuously for the remainder of the experiment. At 2 (○), 4 (●) and 5.5 (●) h, the sensitivity of the nerve to butanol was again determined. The progressive increase in the efficacy of butanol can be clearly seen.

had a small depressant action on the action potential at concentrations up to 5 mM, with little effect on conduction velocity. Similar results have been obtained with the frog sciatic nerve. These results indicate that the local anaesthetic actions of EMD-21657, either alone or with hexanol, cannot be correlated simply with their effects on the physical properties of lipid bilayers.

Temperature effects and unitary hypotheses of local anaesthesia

The interpretation of the effects of temperature on conduction and local anaesthetic action is complicated. The functional properties of the nerve are known to vary with temperature and therefore the sensitivity to perturbation by anaesthetics will be

difficult to predict. Cooling the LOT from 38 to 28 °C in the absence of any anaesthetics slowed conduction without a significant reduction in the amplitude of the compound action potential or increase in its threshold. Further cooling slowed conduction still more, decreased the amplitude of the action potential and increased the threshold until at 13–15 °C, cold block occurred. At the same time, the partition coefficient of the anaesthetic in excitable membranes is likely to be changing. For example, the partition of butanol and benzyl alcohol into various lipid bilayers and erythrocyte membranes decreases with cooling (Fig. 6); this is a further complication in predicting how anaesthetic potency will vary with temperature. However, in spite of these complications, the changes in potency of these two local anaesthetics with temperature are inconsistent with any simple unitary hypothesis based on perturbations of the lipid bilayer; suggestions raised include:

(1) Fluidity changes. The fluidity of the lipid bilayer decreases with decreasing temperature and it might be expected that the anaesthetic potency of both alcohols would also decrease, because their fluidising effect would be antagonised by the reduction in temperature (see Fig. 1), and because their concentrations in the membrane would also be reduced. In fact, the potency of benzyl alcohol as a nerve-blocking agent increases with cooling, so that 10 mM benzyl alcohol attenuated the LOT compound action potential by 30–40% at 37 °C, by 60–75% at 27 °C, and by 80–90% at 20 °C (Fig. 6), whereas the potency of butanol decreased slightly over the same temperature range. These marked differences in the effects of butanol and benzyl alcohol with temperature cannot be rationalised simply in terms of an unitary hypothesis based on fluidity perturbations of the lipid bilayer.

(2) Phase separations. Two recent hypotheses propose that anaesthetics act by fluidising regions of phospholipid which are normally in the liquid-crystalline state. Trudell¹⁷ proposed that when this occurs in the lipid bilayer, the lateral pressure the lipid exerts on the target membrane proteins increases and results in the inhibition of their functions and hence anaesthesia. Alternatively, Lee¹⁸ proposed that the proteins responsible for nerve conduction are maintained in an optimal conformation by an annulus of rigid lipid molecules (similar to a liquid-crystalline state) and that anaesthetics fluidise this lipid to cause the protein to relax into an inactive conformation. Both hypotheses predict that cooling a nerve blocked by an anaesthetic should restore conduction. Furthermore, as nerve block is produced by concentrations of alkanols that decrease the *T_i* of phospholipid bilayers by 3–5 °C, Lee's hypothesis predicts that cooling a nerve by that amount should restore nerve conduction. Both 50 mM butanol and 15 mM benzyl alcohol block conduction in the LOT and produce a decrease in the *T_i* of phospholipid bilayers of about 3 °C. However, in neither case does cooling by this amount restore nerve conduction (see Fig. 6 and above). Thus,

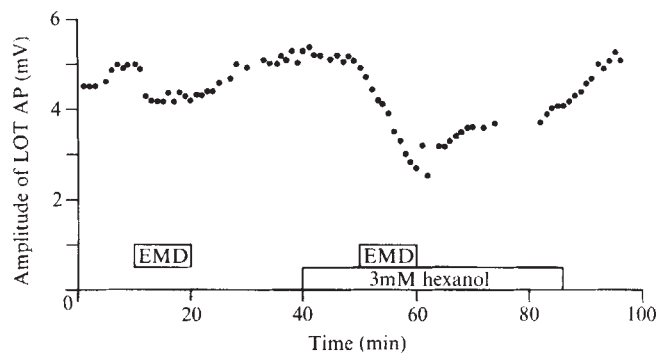


Fig. 5 Interaction between EMD-21657 and hexanol on the amplitude of the LOT compound action potential. EMD-21657 (5 mM) has a slight depressant action but sharply potentiates the action of 3 mM hexanol (3 mM hexanol alone depressed the LOT action potential by 15% in this preparation after 20 min application).

the simplest forms of these two unitary hypotheses are not supported experimentally, and in our view, the wide range of temperatures over which conduction occurs in the LOT and other nerves is a clear indication that perturbation of a critically balanced phase separation is not a likely mechanism of local anaesthetic action. In particular, it is significant that the LOT will conduct at temperatures well above the physiological temperature of 38 °C, at least up to 42 °C, when the annular lipids would be considerably fluidised.

(3) Bilayer thickness. The present data are clearly not directly concerned with the question of changes in bilayer thickness as a mechanism of anaesthetic action, as proposed by Ashcroft *et al.*^{19,20} and Haydon *et al.*^{15,20}. However, both butanol and benzyl alcohol would be expected to decrease bilayer thickness as they expand the area of red cell membranes²² and decrease the gel-liquid phase transition temperatures of artificial phospholipid bilayers²³. This view is consistent with the evidence that anaesthetics decrease the thickness of phospholipid monolayers as their area expands^{24,25}, but contrasts with the evidence from capacitance measurements on black lipid membranes^{19–21}. Even so, we would expect benzyl alcohol and butanol to have similar

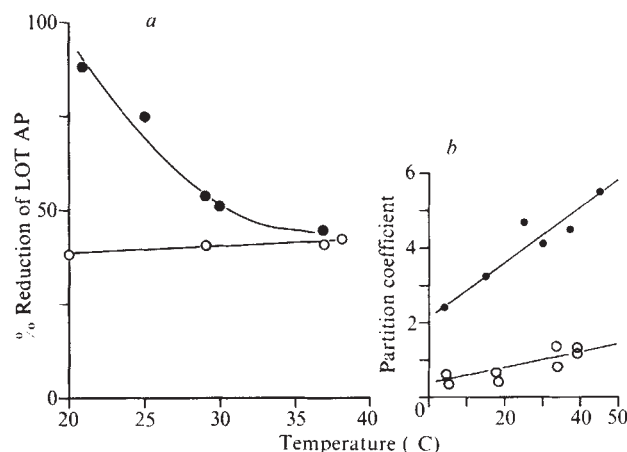


Fig. 6 Effect of temperature on the local anaesthetic actions of butanol and benzyl alcohol and their membrane-buffer partition coefficients. *a*, The results of two experiments are shown. 10 mM benzyl alcohol (●) and 40 mM butanol (○) each reduce the amplitude of the LOT compound action potential by about 40% at 37 °C. The action of 10 mM benzyl alcohol is potentiated by cooling whereas that of 40 mM butanol is not. *b*, Temperature dependence of the membrane-buffer partition coefficients of benzyl alcohol (●) at 30 mM and butanol (○) at 50 mM in liposomes prepared from red cell lipid and phosphatidylcholine: cholesterol films containing 10% phosphatidylserine (cholesterol: phospholipid ratio 1:2), respectively. (Data for benzyl alcohol from ref. 40.) The partition coefficients for butanol were determined using equilibrium dialysis with estimation of butanol by gas chromatography.

effects on bilayer thickness and that their potency as local anaesthetics would vary with temperature in the same direction if their primary action were to be on membrane thickness. The fact that this is not observed experimentally suggests that any changes in bilayer thickness that are produced by alcohols will not correlate directly with their local anaesthetic action.

A degenerate protein perturbation hypothesis

The results presented here show that it is very unlikely that there is a correlation between anaesthetic effects and any of the physical properties of lipid bilayers that have been proposed as the basis of unitary mechanisms of local anaesthetic action. We must therefore conclude that anaesthetics do not generally act by perturbing the lipid bilayer regions of membranes and that

direct interactions between the anaesthetic molecules and target membrane proteins are responsible for local anaesthesia. However, we do not imply that all anaesthetics act by binding to a specific site on one or more of the functional membrane proteins; this is extremely unlikely. Instead, we suppose that, for example, small molecules such as halothane or cyclopropane are distributed in one set of hydrophobic sites of appropriate dimensions within the protein, and that larger molecules such as barbiturates, steroids and aromatic amines, which show varying degrees of stereoselectivity^{26–29}, will bind to distinct sets of sites on the target membrane proteins.

There is some experimental evidence for selectivity in the binding sites of amine local anaesthetics for their action on nerve conduction. For example, their inhibition of the sodium current depends on both the rate of stimulation³⁰ and the size of the voltage step applied to nerves under voltage clamp³¹. This suggests that they bind to a precise part of the sodium conductance mechanism which must possess the appropriate conformation for binding to occur and which may be located within the channel itself. Finally, we suggest that any molecule with structural features similar to those of phospholipids, such as dodecanol, will compete for the hydrophobic regions of membrane proteins which normally interact with the chains of membrane phospholipids, and that replacing the annular lipids³² with anaesthetic molecules interferes with the normal lipid-protein interactions that are required for the maintenance of function³³. The evidence for this type of 'annular displacement mechanism' will be presented elsewhere³⁴. Thus, we propose that the binding of anaesthetics to many different sets of sites on the target proteins can inhibit their function and that the mechanism of anaesthetic action is therefore highly degenerate at the molecular level. We term this the degenerate perturbation hypothesis^{6,35}.

Received 23 August; accepted 30 October 1978.

- Seeman, P. *Pharmac. Rev.* **24**, 583–655 (1972).
- Miller, K. W., Paton, W. D. M., Smith, E. B. & Smith, R. A. *Anesthesiology* **36**, 339–351 (1972).
- Hill, M. W. in *Molecular Mechanisms in General Anaesthesia*, ch. 8 (Churchill-Livingstone, London, 1974).
- Seeman, P. & Roth, S. *Biochim. biophys. Acta* **255**, 171–177 (1972).
- Metcalfe, J. C., Seeman, P. & Burgen, A. S. V. *Molec. Pharmac.* **4**, 87–95 (1968).
- Metcalfe, J. C., Hoult, J. R. S. & Colley, C. M. in *Molecular Mechanisms in General Anaesthesia*, ch. 9 (Churchill-Livingstone, London, 1974).
- Chacko, G. K., Villegas, G. M., Barnola, F. V., Villegas, R. & Goldman, D. E. *Biochim. biophys. Acta* **443**, 19–32 (1976).
- Korn, E. D. *Science* **153**, 1491–1498 (1966).
- Hui, F. K. & Barton, P. G. *Biochim. biophys. Acta* **296**, 510–517 (1973).
- Lee, A. G. *Biochemistry* **16**, 835–841 (1977).
- Price, J. L. & Sprich, W. W. *J. comp. Neurol.* **162**, 321–336 (1975).
- Hesketh, T. R., Keightley, C. A., Metcalfe, J. C. & Richards, C. D. *J. Physiol., Lond.* **278**, 5P–6P (1978).
- Richards, C. D. & Sercombe, R. *J. Physiol., Lond.* **197**, 667–683 (1968).
- Armstrong, C. M. & Binstock, L. J. *gen. Physiol.* **48**, 265–277 (1964).
- Haydon, D. A., Hendry, B. M., Levinson, S. R. & Requena, J. *Biochim. biophys. Acta* **470**, 17–34 (1977).
- Cadwell, P. C. & Keynes, R. D. *J. Physiol., Lond.* **148**, 8P–9P (1959).
- Trudell, J. R. *Anesthesiology* **46**, 5–10 (1977).
- Lee, A. G. *Nature* **262**, 545–548 (1976).
- Ashcroft, R. G., Coster, H. G. L. & Smith, J. R. *Nature* **269**, 819–820 (1977).
- Ashcroft, R. G., Coster, H. G. L. & Smith, J. R. *Biochim. biophys. Acta* **469**, 13–22 (1977).
- Haydon, D. A., Hendry, B. M., Levinson, S. R. & Requena, J. *Nature* **268**, 356–358 (1977).
- Roth, S. & Seeman, P. *Biochim. biophys. Acta* **255**, 190–198 (1972).
- Colley, C. M. & Metcalfe, J. C. *FEBS Lett.* **24**, 241–246 (1972).
- Ueda, I., Shieh, D. D. & Eyring, H. in *Molecular Mechanisms of Anesthesia*, 291–305 (Raven, New York, 1975).
- Trudell, J. R. *Biochim. biophys. Acta* **470**, 509–510 (1977).
- Wahlström, G., Bush, H. & Buzello, T. W. *Acta pharmac. tox.* **28**, 493–498 (1970).
- Phillips, G. H. in *Molecular Mechanisms in General Anesthesia*, ch. 3 (Churchill-Livingstone, London, 1974).
- Richards, C. D. & Hesketh, T. R. *Nature* **256**, 179–182 (1975).
- Åkerman, S. B. A., Camougis, G. & Sandberg, R. V. *Eur. J. Pharmac.* **8**, 337–347 (1969).
- Hille, B., Courtney, K. & Dum, R. in *Molecular Mechanisms of Anesthesia*, 13–24 (Raven, New York, 1975).
- Strichartz, G. J. *gen. Physiol.* **62**, 37–57 (1973).
- Warren, G. B., Toon, P. A., Birdsall, N. J. M., Lee, A. G. & Metcalfe, J. C. *Biochemistry* **13**, 5501–5507 (1974).
- Warren, G. B., Houslay, M. D., Metcalfe, J. C. & Birdsall, N. J. M. *Nature* **255**, 684–687 (1975).
- Keightley, C. A. *et al.*, in preparation.
- Richards, C. D. *Int. Rev. Biochem.* **19**, ch. 6 (1978).
- Robles, E. C. & Van den Berg, D. *Biochim. biophys. Acta* **187**, 520–526 (1969).
- Hubbell, W. L. & McConnell, H. M. *J. Am. chem. Soc.* **93**, 314–326 (1971).
- Richards, C. D., Russell, W. J. & Smajic, J. C. *J. Physiol., Lond.* **248**, 121–142 (1975).
- Richards, C. D. & Tegg, W. J. B. *Br. J. Pharmac.* **59**, 526P (1977).
- Colley, C. M. thesis, Univ. Cambridge (1971).
- Saletu, B., Grünberger, J., Saletu, M., Mader, R. & Volvaiva, J. *Int. Pharmacopsychiat.* **13**, 177–192 (1978).

Multiplicity of germline genes specifying a group of related mouse κ chains with implications for the generation of immunoglobulin diversity

Oscar Valbuena, Kenneth B. Marcu, Martin Weigert & Robert P. Perry

Institute for Cancer Research, the Fox Chase Cancer Center, Philadelphia, Pennsylvania 19111

The number of germline genes specifying the variable sequences of the $V_{\kappa}21$ group of light chains was determined by saturation hybridisation analysis to be four to six. This gene multiplicity is considerably less than the total variability in the group, indicating that κ -chain diversity is generated by somatic mutations in both framework and complementarity-determining (hypervariable) regions.

THE extensive repertoire of immunoglobulins can be accounted for in two ways: multiplicity of germline variable-region genes (V_L and V_H) and somatic diversification of V genes. In the case of mouse λ light chains, the pattern of variability is simple¹. Nucleic acid hybridisation and gene-sequencing studies indicate that variability arises somatically by mutation of one germline V_{λ} gene²⁻⁴. In contrast, for κ chains, which constitute the bulk of the light chains in the mouse, the pattern of variability is extensive, and can be explained at least in part by multiple germline genes. The V_{κ} regions fall into various subgroups that are defined by characteristic amino acid residues and shared sequence insertions or deletions⁵. As multiple, identical somatic mutations would be required to produce V_{κ} regions of two or more subgroups from a single V_{κ} gene, it has been inferred that different subgroups are coded for by different germline genes. The idea of multiple germline V_{κ} genes is consistent with the results of nucleic acid hybridisation studies, which indicate that the messenger RNAs coding for unrelated κ chains have little or no homology in their V-region sequences, and that each of these different species of κ -mRNA seems to be coded for by 2-5 genes per haploid genome⁶⁻¹².

In this study we assess the extent to which somatic mutation might also contribute to V_{κ} diversity. We have approached this problem by determining the germline gene multiplicity for a group of related κ chains. This group, called $V_{\kappa}21$ according to the recent classification of mouse V_{κ} sequences¹³, is composed of 30 examples of κ chains produced by independently induced plasmacytomas of either NZB or BALB/c mice (see accompanying paper¹⁴ and ref. 15). The amino-terminal sequences (positions 1-23) of these κ chains are identical except for a single substitution. Throughout the rest of the V region (positions 24-112) there is considerable diversity. Among the 22 NZB examples of $V_{\kappa}21$, 20 unique sequences have been found with differences both in the framework and complementarity-determining (hypervariable) regions.

Our estimate of the number of genes coding for selected $V_{\kappa}21$ -mRNAs indicates that there are only four to six germline

V genes for the entire $V_{\kappa}21$ group, a value far too small to account for the diversity of this group. This provides a compelling argument for the importance of somatic processes in the generation of κ -chain diversity. Moreover, these results indicate that such somatic processes can result in extensive substitutions in the framework region as well as in the complementarity-determining regions.

Determination of the nucleotide sequence homology among κ -mRNAs

The number of $V_{\kappa}21$ genes can be determined by nucleic acid hybridisation measurements with probes containing V-region sequences of selected $V_{\kappa}21$ -mRNAs. To interpret such measurements correctly, it is necessary to know the capacity of a particular V probe to anneal to various V-region sequences within and outside the $V_{\kappa}21$ group. For this reason, we first examined the cross-hybridisation between a labelled complementary DNA representing one member (MOPC 321) with three other selected mRNAs of the group, those coding for PC 3741, PC 7043 and PC 2880 κ chains. These four light chains

Table 1 Cross-hybridisation of full-length MOPC 321 κ -cDNA with κ -mRNAs

mRNA	% protection of full-length MOPC 321 κ -cDNA	% of amino acid homology in the variable region
κ (MOPC 321)	97	100
κ (PC 3741)	98	>94
κ (PC 7043)	98	85
κ (PC 2880)	92	80
κ (MPC 11)	64	~50
C_{κ} (MPC 11)	68	0
λ (MOPC H2020)	3	
α (MOPC H2020)	3	
None	2	

Hybridisation reactions were carried out at 70 °C in 50- μ l reaction mixtures¹⁸ containing 10 ng of mRNA and 0.4-0.5 ng of ³H-cDNA. All mRNAs except C_{κ} were purified by two cycles of sucrose gradient sedimentation. The C_{κ} -mRNA was purified by sucrose gradient sedimentation and electrophoresis on a formamide-polyacrylamide gel. The reactions were carried out to a C_{κ} of 5×10^{-2} mol l⁻¹ s and assayed with S₁ nuclease as described previously¹⁸. The percentages of protection are corrected for 2% self-annealing of the probe at the longest incubation period. The percentages of amino-acid homology in the variable regions refer to the first 98 N-terminal amino acids¹⁴.

differ from each other in the framework as well as the complementarity-determining regions¹⁴. Relative to MOPC 321 κ chain, the amino-acid sequence homologies in the variable regions of the PC 3741, PC 7043 and PC 2880 κ chains are >94%, 85% and 80%, respectively (Table 1). The variable sequences of MOPC 321 and PC 2880 are among the most distantly related in the $V_{\kappa}21$ group. For an example of an unrelated V_{κ} sequence we used the mRNA coding for MPC 11 κ chain, a member of the $V_{\kappa}19$ group, which differs from the MOPC 321 κ chain at 10 of the first 23 N-terminal residues and about half of the total variable-region positions^{5,16}. As a control for null hybridisation of V-region sequences we used an mRNA

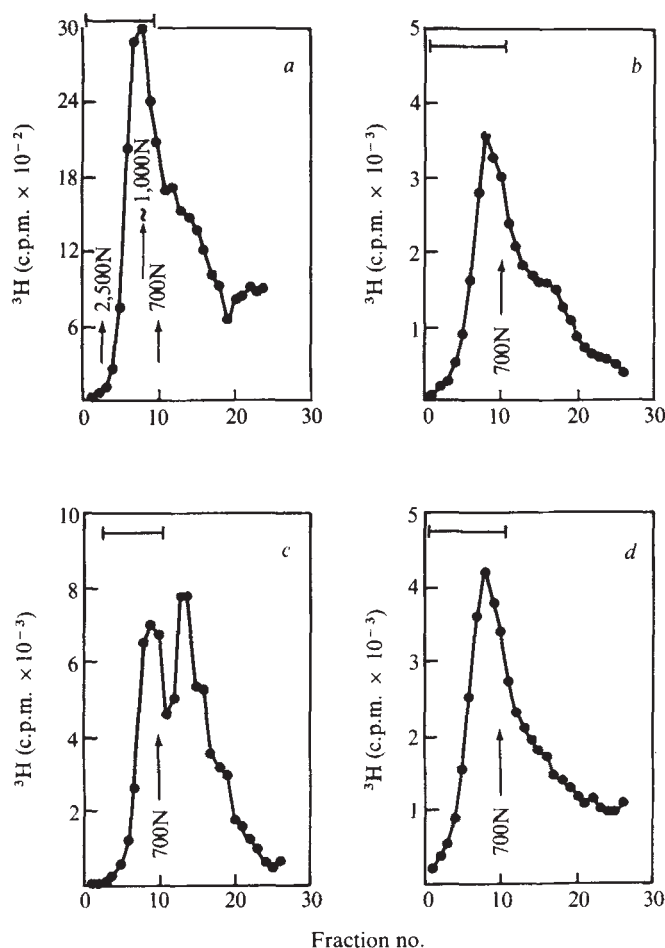


Fig. 1 Isolation of full-length κ -cDNA in alkaline sucrose gradients. 300–500 ng of ^3H -cDNAs were sedimented through 5–20% alkaline sucrose gradients in a SW50.1 rotor, at 36,000 r.p.m., at 22 °C for 16 h. Fractions of 200 μl were collected and the radioactivity determined by counting 5–10- μl aliquots in a scintillation counter. Arrows indicate the position of polyoma DNA fragment markers (2,500 and 700 nucleotides (N) long) analysed in parallel gradients. Panels a–d show the profiles of the total cDNA preparations synthesised from MOPC 321, PC 2880, PC 3741 and MPC 11 κ -mRNAs, respectively. The fractions corresponding to full-length cDNAs (horizontal bars) were pooled, ethanol precipitated, and used for subsequent hybridisation studies.

derived from MPC 11 cells (C_{κ} -mRNA), which lacks the sequences coding for the variable region of the κ chain¹⁷.

The mRNAs used in these experiments were extracted from membrane-bound polyribosomes and purified by oligo dT-cellulose chromatography, sedimentation through successive sucrose gradients and, in some cases, electrophoresis in

formamide-polyacrylamide gels^{18–20}. Their purity, evaluated by kinetic complexity analysis or oligonucleotide fingerprinting and 5'-terminal analysis, ranged from 75% to close to 100% (refs 18–20). The ^3H -cDNAs (10 c.p.m. per pg) were synthesised with avian myeloblastosis virus reverse transcriptase and oligo dT primer in conditions which yield a high proportion of full-length product. Analyses of the cDNAs on alkaline sucrose gradients showed distinct peaks at ~1,000 nucleotides and a trailing of shorter, incomplete transcripts (Fig. 1). To maximise the equivalent representation of 3' (C region) and 5' (V region) sequences we used only the cDNA in the fastest sedimenting (1,000-nucleotide peak) fractions. The probes were further purified by back hybridisation to their mRNA templates in conditions in which only the most abundant sequences would anneal¹⁸. When rehybridised to the mRNA templates at C_{t} values of about 10^{-2} or greater, 90–100% of the purified cDNA formed S_1 nuclease-resistant duplex.

When full-length MOPC 321 κ -cDNA was hybridised to κ -mRNAs of MOPC 321, PC 3741, PC 7043 and PC 2880, 97%, 98%, 98% and 92% of the probe, respectively, was incorporated into S_1 nuclease-resistant hybrid (Table 1). The high proportions of MOPC 321 κ -cDNA protected from S_1 nuclease digestion by the heterologous κ -mRNAs indicate that the four mRNAs have a high degree of nucleotide sequence homology. In contrast, the MOPC 321 κ -cDNA probe was not protected by MOPC H2020 λ - or MOPC H2020 α -mRNAs, thus verifying the high specificity of the cDNA preparation. When the MOPC 321 κ -cDNA was hybridised to MPC 11 κ -mRNA or to C_{κ} -mRNA, approximately 65% of the probe was protected. A similar result was obtained when a full-length MPC 11 κ -cDNA probe was annealed to MOPC 321 κ -mRNA (data not shown). This extent of hybridisation is attributable to a common 3' portion of the mRNA molecules, which contains the coding sequences for the constant regions of the immunoglobulin chains. Assuming that the 3' untranslated sequences of the κ -mRNAs are identical or very similar for different κ -mRNAs, it seems reasonable to conclude that the additional 30–35% protection exhibited when the other members of the $V_{\kappa}21$ group of mRNAs were hybridised to the MOPC 321 κ -cDNA probe represents hybridisation to common variable nucleotide sequences located on the 5' portions of the mRNA molecules. All the hybrids formed in this analysis had melting temperatures above 90 °C (ref. 19), as one would expect for relatively well matched DNA-RNA duplexes.

The above results indicate that the mRNAs specifying the various members of the $V_{\kappa}21$ group have sufficiently similar V-region sequences for their cDNA probes to hybridise effectively to all the $V_{\kappa}21$ genes in the reaction conditions used, particularly if complications of sister-strand competition are avoided (see below). On the other hand, a probe representing an unrelated V sequence, such as that corresponding to the variable portion of MPC 11 κ -mRNA, would not be expected to hybridise to a gene of the $V_{\kappa}21$ group, and conversely, a $V_{\kappa}21$ probe would not be expected to recognise the gene coding for the MPC 11 κ -variable sequence. Previous measurements of cross hybridisation between other V_{κ} sequences have led to similar conclusions^{2,21}.

The high degree of homology among $V_{\kappa}21$ -mRNAs suggests that nucleotide substitutions closely parallel amino-acid substitutions within the $V_{\kappa}21$ group. The κ chains of MOPC 321 and PC 2880, the two most distant $V_{\kappa}21$ members among those we have investigated, have different amino acids at 20 positions within the first 107 residues from the N-terminal end, which could be accounted for by a minimum of 23 nucleotide substitutions¹⁴. Assuming no other substitutions in the variable region and identity in the constant and untranslated regions of the mRNA, the minimum nucleotide divergence would be approximately 23/1,000 or 2.3% of the total mRNA sequence. Such a divergence might be manifested by a slight (2–4 °C) difference between the melting temperatures of homologous and heterologous hybrids¹², but it would not be readily detectable by the hybridisation assays used in these experiments. The absence

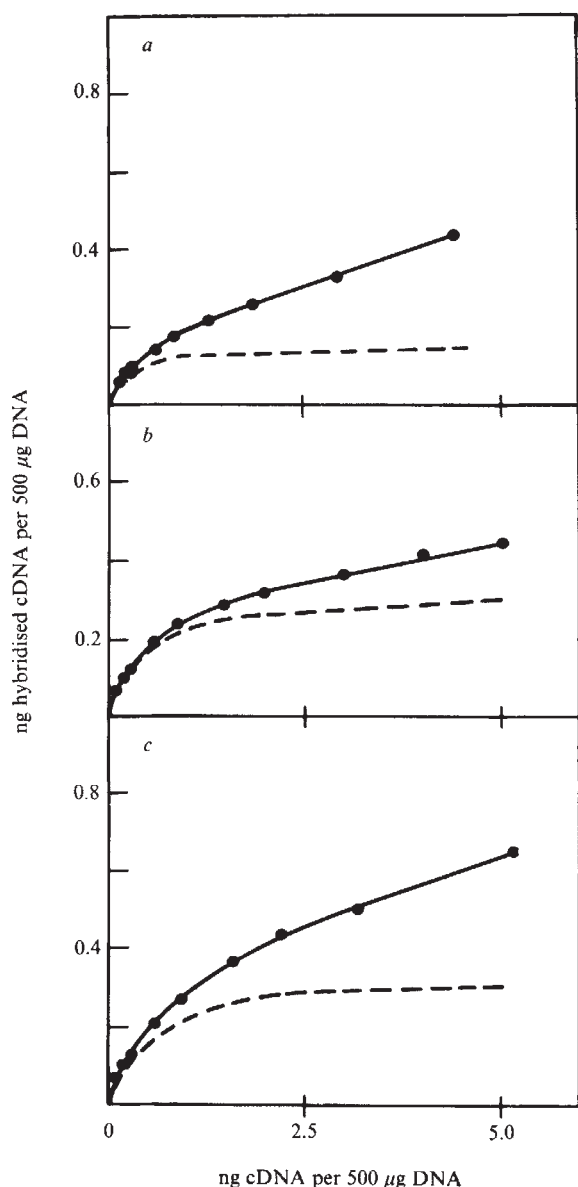


Fig. 2 Saturation of mouse embryo DNA with full-length κ -cDNAs. For each point 510 μg of NZB mouse embryo DNA, cleaved into fragments of about 500 nucleotides average chain length¹⁹, was hybridised at 68 °C to a C_0t of $6 \times 10^4 \text{ mol l}^{-1} \text{ s}$ with an increasing amount of ^3H -cDNA probe (0.1–5.0 ng, $10^4 \text{ c.p.m. per ng}$) in a total volume of 50 μl containing 0.6 M NaCl, 40 mM Tris-Cl, pH 7.2, 2 mM ethylene diamine tetraacetate, 1 mg ml^{-1} yeast RNA and 0.2 $\mu\text{g ml}^{-1}$ poly A. At the beginning of the hybridisation incubation the mixture was heated at 100 °C for 10 min. The hybrids were assayed after treatment with S_1 nuclease¹⁸. *a*, *b* and *c* show saturation data for C_κ , MOPC 321 and PC 2880 κ -cDNAs, respectively. Solid lines represent the hybridised cDNA (corrected for 2% self-annealing of the probes) and broken lines represent hybridised cDNA after correction for contamination (see text). The C_κ -cDNA was prepared by incubating a full-length MOPC 3741 κ -cDNA with the unrelated MPC 11 κ -mRNA and purifying the hybrid by S_1 nuclease digestion. This cDNA consists of constant-region sequences plus untranslated sequences that are common to the PC 3741 and MPC 11 κ -mRNAs.

of cross-hybridisation between the variable sequences of MOPC 321 and MPC 11 κ -mRNAs suggests that between the two sets of corresponding genes there are, in addition to the divergence leading to amino-acid substitutions, substantial differences due to neutral mutations^{22–25}. Such a situation is consistent with an evolutionary model in which there is early

divergence of the ancestral genes for MOPC 321 and MPC 11 V_κ sequences and a more recent expansion of these individual gene sets.

Estimation of gene number

The gene multiplicity was determined by saturation experiments, in which a fixed amount of mouse embryo DNA was hybridised with increasing amounts of full-length κ -cDNA probes. When the probe is in large excess relative to the corresponding gene sequence, the elimination of reactive loci due to the annealing of (–) DNA sister strands is negligible, and one may expect that all the relevant genes will be duplexed with probe. This feature is particularly advantageous for measuring the aggregate number of closely related genes because it minimises the possibility of related but non-identical sequences being out-competed by perfectly matched (–) DNA strands. As this effect would be magnified when gene sequences are in excess, the validity of immunoglobulin gene copy numbers determined by reassociation kinetic analysis has been questioned^{2,26,27}.

When 510 μg of embryo DNA was incubated to a C_0t of $6 \times 10^4 \text{ mol l}^{-1} \text{ s}$ with increasing amounts of probe, data such as those shown in Fig. 2 were obtained. The saturation curves were apparently biphasic and did not exhibit true saturation plateaux. The biphasic nature was more clearly revealed when double-reciprocal plots of the saturation data were made (Fig. 3). All the curves consisted of a linear component representing the saturation behaviour of the predominant cDNA species, and a deviation from linearity at high cDNA inputs, which is most readily interpreted as representing the hybridisation of a small, but diverse fraction of contaminating non- κ -chain sequences, each of which is present in subsaturating amounts. This interpretation was substantiated by several facts. First, when the input of MOPC 321 κ -cDNA was 12.5 ng (2.5 times the highest input used in Fig. 2), the saturation curve still failed to approach a plateau, and, in fact, continued to rise with constant slope. Second, the levels of putative contaminating non- κ -chain sequences in the MOPC 321 and MPC 11 cDNA probes calculated from the reciprocal plot analysis (3% and 9%, respectively, see below) are roughly proportional to the levels of impurities in the respective mRNA preparations, as judged from kinetic complexity analyses¹⁸. Third, when saturation experiments were carried out with homogeneous κ -mRNA sequences prepared from cloned DNA, true saturation plateaux and completely linear reciprocal plots were observed (Fig. 4). Moreover, the gene numbers estimated from these saturation values did not differ significantly from the corresponding numbers obtained with enzymatically synthesised probes when corrections for the contaminating components had been made (see below).

To calculate the saturation values corresponding to each κ -chain species, we empirically assigned a percentage of contamination to each probe, which if corrected for, would result in a completely linear curve in the reciprocal plot (Fig. 3). The levels of contamination estimated by trial and error were 2.9%, 6.5%, 6.5% and 9.0% for MOPC 321 κ -cDNA, PC 2880 κ -cDNA, C_κ -cDNA and MPC 11 κ -cDNA, respectively. When these proportions of nonsaturating components were subtracted from the hybridisation values, reasonably good plateaux were obtained (broken curves, Fig. 2). The saturation values (H_∞) were then calculated from the corrected ordinate intercepts of the double-reciprocal plots (Fig. 3). The high precision of the saturation data enabled us to calculate the H_∞ values with a maximum probable error of $\pm 10\%$. These values, together with the molecular weight of the mouse haploid genome and the molecular weights of the C_κ and V_κ segments, were used to calculate the number of gene copies (Table 2). One set of gene numbers was calculated using 650 nucleotides and 350 nucleotides, respectively, for the C_κ and V_κ molecular weights, based on the fact that about 65% of full-length MOPC 321 κ -cDNA (1,000 nucleotides) is protected by MPC 11 κ -mRNA or C_κ -

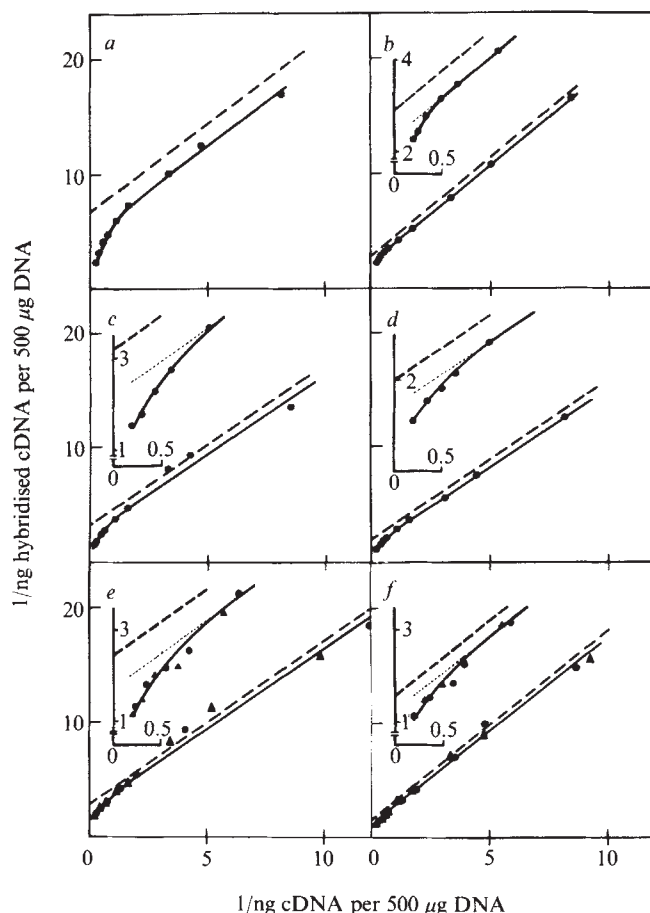


Fig. 3 Double-reciprocal plots of the cDNA saturation data. Solid lines represent hybridised cDNA (2% self-annealing subtracted). Broken lines represent hybridised cDNA after correction for contamination of probe by non- κ -chain sequences. The intercept and slope of the broken lines were estimated by the least squares method. *a-d* correspond to C_{κ} , κ (MOPC 321), κ (PC 2880) and κ (MPC 11) cDNA saturation reactions, respectively. *e* and *f* correspond to equimolar mixtures of κ (MOPC 321) + κ (PC 2880) and κ (MOPC 321) + κ (MPC 11) cDNAs, respectively. $\blacktriangle$ and $\bullet$ refer to two independent experiments. Contaminant correction for the mixtures was made using an average for the individual probes. The portions of the curves with the higher cDNA inputs are enlarged in the insets.

mRNA sequences (Table 1). An alternative set of gene numbers was calculated using 521 and 479 nucleotides, respectively, for the C_{κ} and V_{κ} segments, based on current ideas of the structure of a κ -mRNA²⁸. With this information, the calculated gene multiplicities per haploid genome are as follows: 2–3 genes for C_{κ} ; 4–6 V_{κ} genes for MOPC 321; 4–5 V_{κ} genes for PC 2880; 8–11 V_{κ} genes for MPC-11. It is evident that all the tested V_{κ} genes are in a higher number of copies than the C_{κ} gene.

When a 1:1 mixture of MOPC 321 and MPC 11 κ -cDNAs was used, the measured number of V_{κ} genes is in good agreement with the sum of the individual gene reiteration values. On the other hand, when a mixture of MOPC 321 + PC 2880 κ -cDNAs was used, no additive hybridisation was observed; the measured number of V_{κ} genes for this mixture is not significantly different from either of the individual V_{κ} -gene reiteration values for MOPC 321 or PC 2880. These results support the inferences made on the basis of the cross-hybridisation experiments. The additivity exhibited by MOPC 321 and MPC 11 V_{κ} sequences demonstrates that the two probes are hybridising to separate genetic loci, thus implying two different

sets of genes. The non-additivity displayed by MOPC 321 and PC 2880 V_{κ} sequences indicates that both probes are hybridising to the same 4–6 V_{κ} 21-related loci.

The saturation data with the cloned DNA sequences were also used to calculate gene numbers. In this case, estimates of the sizes of the C_{κ} and V_{κ} segments were made from measurements of the amounts of cloned MPC 11 κ -DNA sequence protected by MOPC 321 κ -mRNA (70%) and MPC 11 κ -mRNA (100%). As the cloned MPC 11- κ -sequence was calculated to

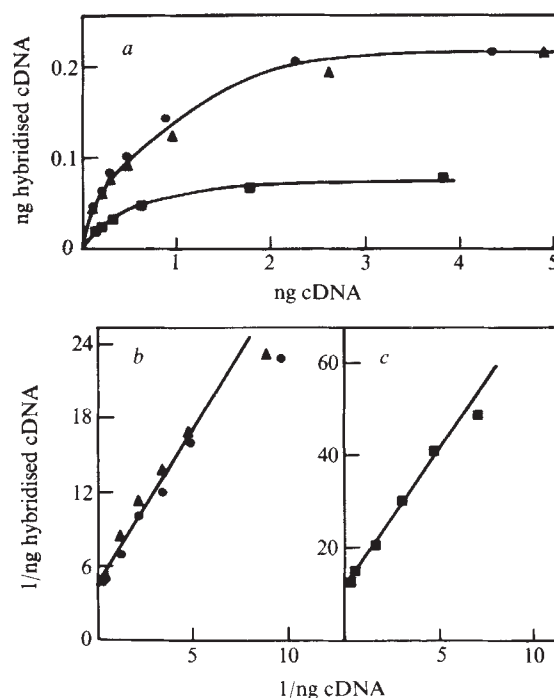


Fig. 4 Saturation-hybridisation reactions with cloned MPC 11 ¹²⁵I- κ -cDNAs. The cloned MPC 11 κ -mRNA sequence is carried as a 700-nucleotide insert in the plasmid pMB9. The recombinant DNA was iodinated to a specific activity of 60 c.p.m. per pg, denatured and hybridised in 80% formamide for 1 h at 50°C to an excess of MPC 11 κ -mRNA. The hybridisation mixture was treated with S_1 nuclease, and the ¹²⁵I- κ -DNA sequences recovered and submitted to a second round of purification by hybridisation with either MPC 11 κ -mRNA or MOPC 321 κ -mRNA and subsequent treatment with S_1 nuclease¹⁹. The resultant homogeneous probes represent C-plus-V-region sequences or predominantly C-region sequences, respectively. 0.448 mg of mouse embryo DNA was hybridised to increasing amounts of cloned probes (0.1–5.0 μ g) in a total volume of 50 μ l. The reactions were incubated and assayed as described in Fig. 2. The final C_0t value in this set of reactions was 5.4×10^4 mol l⁻¹ s. *a*, Saturation curves for MPC 11 κ -DNA ($\blacktriangle$ and $\bullet$ refer to two independent experiments) and for C_{κ} -DNA ($\blacksquare$). The hybridisation values were corrected for self-annealing (6% for MPC 11 κ -DNA and 15% for C_{κ} -DNA). *b*, *c*, Double-reciprocal plots of the MPC 11 κ -DNA and C_{κ} -DNA saturation curves, respectively.

be 700 nucleotides (U. Schibler, personal communication), 70% or 490 nucleotides of the probe represent C_{κ} sequences and 30% or 210 nucleotides represent V_{κ} sequences. With these parameters, gene copy numbers of 2 and 8 per haploid genome for C_{κ} and MPC 11-related V_{κ} genes were calculated (Table 2). These values are in excellent agreement with the values obtained with enzymatically synthesised cDNAs after corrections for the non- κ -chain contaminants had been made, thus indicating that such corrections do not contribute significant errors to our estimates of gene number.

The basis of $V_{\kappa}21$ diversity

The $V_{\kappa}21$ group of immunoglobulins is apparently represented in germline DNA by four to six genes specifying the variable-region sequences. These same four to six genes are recognised by sequence probes corresponding to any member of the group, even variants such as MOPC 321 and PC 2880 which must differ, as inferred from their amino acid sequences, by at least 7% of the nucleotides in the V-region sequence. In contrast, amino-acid sequence studies^{14,15} have revealed extensive diversity in the $V_{\kappa}21$ chains; if only the NZB $V_{\kappa}21$ sequences are compared (to preclude any contribution of genetic polymorphism), 20 different sequences are found in the region 1–112. From the pattern of variability, $V_{\kappa}21$ may be subdivided into six subgroups, $V_{\kappa}21$ A to F, on the basis of two or more $V_{\kappa}21$ regions sharing at least three residues that distinguish members of a particular set of $V_{\kappa}21$ sequences from any other $V_{\kappa}21$ sequence. The sequences that define $V_{\kappa}21$ A–F extend only to residue 98; the rest of what has conventionally been called V_L gene (to residue 112) shows a distinctive pattern of variability that is independent of the pattern over 1–98. It has been proposed that $V_{\kappa}21$ is coded for by at least six germline genes coding for residues 1–98 or 99. The rest of the V region (99 or 100 to 112) is coded for by a separate set of genes, termed J genes. This view of $V_{\kappa}21$ gene organisation is analogous to that recently proposed for the mouse λ gene^{4,29}.

Table 2 Gene copy numbers estimated from saturation-hybridisation experiments

cDNA	H_{∞}	$n(a)$	$n(b)$	$n(c)$
Constant genes				
C_{κ}	0.15	2.4	3.0	
cloned C_{κ}	0.08			2.0
Variable genes				
κ (MOPC 321)	0.35	6.1	4.4	
κ (PC 2880)	0.31	4.9	3.6	
κ (MPC 11)	0.49	10.6	7.7	
cloned κ (MPC 11)	0.22			8.1
κ (MOPC 321) + κ (PC 2880)	0.34	5.9	4.3	
κ (MOPC 321) + κ (MPC 11)	0.63	14.9	10.8	

H_{∞} is the mass of cDNA (in ng) hybridised at infinite input, obtained from the ordinate intercepts of Fig. 3. The number of genes, n , is calculated assuming that the probes contain C_{κ} and V_{κ} sequences, respectively, comprising: a , 650 and 350 nucleotides; b , 520 and 480 nucleotides; c , 490 and 210 nucleotides. For constant genes, $n = (H_{\infty}^C \times MW_C) / (A \times MW_C)$. For variable genes, $n = (H_{\infty}^{V+C} - H_{\infty}^C) MW_V / (A \times MW_V)$. H_{∞}^C and H_{∞}^{V+C} are the H values for the C and C+V probes, respectively. MW_C , MW_V and MW_{V+C} are the molecular weights of the mouse genome (1.8×10^{12}), the C_{κ} and the V_{κ} sequences, respectively. A is the mass of mouse embryo DNA per reaction (0.45 or 0.51 mg).

The estimate of six germline $V_{\kappa}21$ genes based on shared amino-acid sequences is consistent with our finding that $V_{\kappa}21$ is represented by at most six germline genes. The additional diversity in $V_{\kappa}21$, over and above that due to multiple $V_{\kappa}21$ genes, must be due to somatic diversification. It is plausible that one source of such diversification is joining of different $V_{\kappa}21$ genes with different J genes. So far, amino-acid sequencing data suggest that such a process involves only two gene segments. However, $V_{\kappa}21$ examples might still be found with other identical amino-acid substitutions which would define new subgroups. In this case, additional J-type segments would need to be invoked³⁰.

Extensive diversity is also observed in comparisons of $V_{\kappa}21$ sequences in the region 1–98. Most of the sequence differences occur in the complementarity-determining regions. In this sense

the pattern of variability is similar to that found in mouse V_{κ} , in which variability is confined to the complementarity-determining regions¹. That the observed variability is concentrated in these regions could be due to somatic mutation and selection by antigen of mutants with appropriate specificities. In contrast to V_{κ} , however, variability in the 1–98 region of $V_{\kappa}21$ is also seen outside the complementarity-determining regions, that is, in the framework regions. The 10 different framework sequences seen in the limited survey already exceeds the six that we can account for by different germline genes. That somatic variation is observed in the framework portions could imply that variation in the framework can also serve to modify antibody specificity. Alternatively, somatically derived variation both in the complementarity-determining and framework regions might reflect features of the rate or specificity of the somatic process.

The degree to which somatic diversification contributes to total V_{κ} -chain diversity can be assessed from our method for counting germline genes. We have shown that four to six germline genes code for $V_{\kappa}21$ and about eight code for an unrelated group represented by the MPC 11 κ chain. Similar values have been obtained for the MOPC 21 κ chain, a member of the $V_{\kappa}15$ group, which is unrelated to either the $V_{\kappa}21$ examples or MPC 11 (ref. 12). If each V_{κ} group is coded for by four to eight genes, then the total germline gene content for V_{κ} will be four to eight times the number of such unrelated V_{κ} groups. This latter value has been estimated to be about 50, based on the repeat frequency of identical sequences in the region 1–23 found in a group of randomly chosen NZB κ chains¹. Hence, the total number of V_{κ} germline genes can be estimated to be about 300. If the germline gene content for V_H is correspondingly high, then the potential germline repertoire of antibody specificities could be of the order of 10^5 . This value, although large, still falls short of the estimated repertoire ($>10^6$) that an inbred mouse seems to express²⁶. Hence, somatic diversification, as shown here for $V_{\kappa}21$, must also be responsible for producing most antibody specificities.

This research was supported by a grant from the NSF, grants from the NIH and an appropriation from the Commonwealth of Pennsylvania. O.V. is the recipient of a fellowship from the Venezuelan government. The cloned MPC 11 κ -mRNA sequence was provided by Dr Ueli Schibler.

Received 31 July; accepted 31 October 1978.

- Weigert, M. & Riblet, R. *Cold Spring Harb. Symp. quant. Biol.* **41**, 837–846 (1976).
- Leder, P., Honjo, T., Seidman, J. & Swan, D. *Cold Spring Harb. Symp. quant. Biol.* **41**, 855–862 (1976).
- Tonegawa, S., Hozumi, N., Matthysens, G. & Schuller, R. *Cold Spring Harb. Symp. quant. Biol.* **41**, 877–889 (1976).
- Bernard, O., Hozumi, N. & Tonegawa, S. (in preparation).
- Gray, W. R., Dreyer, W. J. & Hood, L. *Science* **155**, 465–467 (1967).
- Honjo, T., Packman, S., Swan, D., Nau, M. & Leder, P. *Proc. natn. Acad. Sci. U.S.A.* **71**, 3659–3663 (1974).
- Tonegawa, S., Steinberg, C., Dube, S. & Bernardini, A. *Proc. natn. Acad. Sci. U.S.A.* **71**, 4027–4031 (1974).
- Rabbitts, T. H. & Milstein, C. *Eur. J. Biochem.* **52**, 125–133 (1975).
- Schechter, I. *Proc. natn. Acad. Sci. U.S.A.* **72**, 2511–2514 (1975).
- Farace, M. G. *et al. Proc. natn. Acad. Sci. U.S.A.* **73**, 727–731 (1976).
- Legler, M. K. & Cohen, E. P. *Biochemistry* **15**, 4390–4399 (1976).
- Rabbitts, T. H. *Immun. Rev.* **36**, 29–50 (1977).
- Potter, M. *Adv. Immun.* **5**, 141–212 (1977).
- Weigert, M., Gattaitan, L., Loh, E., Schilling, J. & Hood, L. *Nature* **276**, 785–790 (1978).
- McKean, D., Potter, M. & Hood, L. *Biochemistry* **12**, 760–771 (1973).
- Smith, G. P. *Science* **181**, 941–943 (1973).
- Kuehl, W. M., Kaplan, B. A., Scharff, M. D., Nau, M., Honjo, T. & Leder, P. *Cell* **5**, 139–147 (1975).
- Marcu, K. B., Valbuena, O. & Perry, R. P. *Biochemistry* **17**, 1723–1733 (1978).
- Valbuena, O. E. thesis, Univ. Pennsylvania (1978).
- Marcu, K. B., Schibler, U. & Perry, R. P. *J. molec. Biol.* **120**, 381–400 (1978).
- Stavnezer, J. & Bishop, J. M. *Biochemistry* **16**, 4225–4232 (1977).
- Robertson, H. D. & Jeppesen, P. G. N. *J. molec. Biol.* **68**, 417–428 (1972).
- Farquhar, M. N. & McCarthy, B. J. *Biochemistry* **12**, 4113–4122 (1973).
- Leder, P. *et al. Cold Spring Harb. Symp. quant. Biol.* **38**, 753–761 (1973).
- Gummerson, K. S. & Williamson, R. *Nature* **247**, 265–267 (1974).
- Williamson, A. R. A. *Rev. Biochem.* **45**, 467–500 (1976).
- Smith, G. P. *Cold Spring Harb. Symp. quant. Biol.* **41**, 863–875 (1976).
- Milstein, C., Brownlee, G. G., Cartwright, E. M., Jarvis, J. M. & Proudfoot, N. J. *Nature* **252**, 354–358 (1974).
- Brack, C., Hiram, M., Lenhard-Schuller, R. & Tonegawa, S. *Cell* (in the press).
- Kabat, E. A., Wu, T. T. & Bilofsky, H. *Proc. natn. Acad. Sci. U.S.A.* **75**, 2429–2433 (1978).

Rearrangement of genetic information may produce immunoglobulin diversity

M. Weigert & L. Gatmaitan

Institute for Cancer Research, Fox Chase Cancer Center, Philadelphia, Pennsylvania 19111

E. Loh*, J. Schilling & L. Hood

Division of Biology, California Institute of Technology, Pasadena, California 91125

The nearly complete amino-acid sequences of 22 closely related immunoglobulin κ variable (V_κ) regions from the inbred NZB mouse are presented. This group of V_κ regions is encoded by at least six germline V_κ genes. These data also suggest that the mouse κ gene is divided into three segments termed V or variable (residues 1 to 98 or 99), J or joining (residues 99 or 100 to 112) and C or constant (residues 113–219). Tonegawa et al. have recently described a similar J segment for mouse λ chains. Inbred mice contain multiple V_κ and J_κ gene segments. Therefore, different combinations of V and J gene segments may be joined at the DNA level during the differentiation of individual lymphocytes to contribute to antibody diversity.

RECENT structural studies on antibody molecules, mRNAs and genes have contributed substantially to our understanding of antibody diversity. Antibody polypeptides are divided into two regions, an N-terminal variable (V) region and a C-terminal constant (C) region. The organisation of antibody genes seems unusual in two respects: (1) the V and C regions are encoded by separate genes in the germline of each organism^{1,2} and (2) these genes are rearranged with respect to one another during the differentiation of the antibody-producing cell^{3,4}. Variable regions, which encode the antigen-binding sites, are the focus for three theories of antibody diversity. The germline theory contends that antibody variability is encoded by large numbers of germline V genes⁵. The somatic mutation theory postulates that diversity arises by somatic variation of relatively few germline genes⁶. A third model suggests that the rearrangement of antibody gene segments during differentiation may also generate antibody variability^{7–9}. The studies described here on mouse κ chains stress the importance of multiple germline V genes and suggest that antibody diversity may be increased by the combinatorial joining of multiple V and J gene segments.

Diversity in mouse light chains

Myeloma tumours artificially induced in two inbred strains of mice, BALB/c and NZB, have provided homogeneous immunoglobulins, mRNAs and chromosomal DNA for the analysis of the diversity patterns of V regions in the two light-chain immunoglobulin families, λ and κ ¹⁰. The overall pattern of V-region diversity in mouse light chains shows that most of the variability resides in three hypervariable regions. These fold to constitute the light-chain contribution to the walls of the antigen-binding crevice. The remainder of the V region provides a scaffold for the binding site and is termed the framework region^{7,11–13}. The extent and nature of diversity patterns in V_λ and V_κ are quite different.

The sequences of mouse λ chains suggest that somatic muta-

tion may have an important role in generating antibody diversity. Amino acid sequence analyses have established that 12 of 18 V_λ regions studied are identical and that the remaining six differ by one to three amino acid substitutions, all of which occur in one of three hypervariable regions¹⁵. It has been proposed that the 12 identical V_λ regions ($V_{\lambda 0}$) are encoded by a single germline gene and that the variants arise by somatic mutations⁶. This interpretation is supported by nucleic acid hybridisation studies which suggest that V_λ regions are encoded by one or a few germline genes^{16,17}. In addition, the DNA sequence of the putative germline V_λ gene, $V_{\lambda 0}$, has been determined for a V_λ gene isolated from presumably undifferentiated mouse embryo DNA and a variant sequence has been determined on another V_λ gene isolated from myeloma DNA¹⁸. (A 19th λ chain, MOPC 315, seems to differ by about 10% of its V sequence from the other 18 λ chains. This V region is probably encoded by a second V_λ gene.) Thus, most mouse λ chains exhibit a simple pattern of diversity at the protein and nucleic acid levels that is consistent with one germline V gene and variants arising by somatic mutations.

Variability in mouse V_κ regions is extensive in both hypervariable and framework regions^{11,19}. A relatedness tree for the N-terminal 23 residues, a portion of the framework region, of 61 BALB/c and NZB κ chains depicts 55 different sequences (Fig. 1). We will refer to a set of V_κ regions which are very similar over their N-terminal 23 residues as a group. (Potter has suggested that each set of mouse V_κ regions differing by three or fewer residues at their N terminus be given number (that is, group) designations²⁰. By his criterion, 64 BALB/c V_κ sequences fall into 26 groups and 31 NZB V_κ sequences fall into 14 groups.) The complete V-region sequences from distinct groups may differ by up to 50% of their sequence. The genetic interpretation of V-region groups is that each is encoded by at least one distinct germline V gene because multiple, identical somatic mutations would be required to produce the immunoglobulin V regions of two or more groups from a single V_κ gene^{21,22}. This is described as the argument of parallel mutation.

Nucleic acid hybridisation data support the supposition that different groups are encoded by distinct V genes²³. Hence, the minimum number of germline genes in a particular antibody family can be estimated by counting V-region groups. In the mouse κ family, the number of groups has been estimated to be about 50 for either BALB/c or NZB mice¹⁵. If the myeloma data analysed do not constitute a random selection of the normal V_κ genes, the number of groups could be larger. In striking contrast, most λ chains fall into a single group.

The variability of mouse κ chains within a single group, $V_\kappa 21$, does not resemble that of the predominant λ group (for example, with variability only in hypervariable regions). This has been shown for eight BALB/c κ chains in the $V_\kappa 21$ group that have identical N-terminal sequences except for a single substitution. Three pairs of these V_κ regions fall into closely related sets or subgroups, designated $V_\kappa 21A$, $V_\kappa 21B$ and $V_\kappa 21C$, which are defined by multiple shared or subgroup-associated amino acid residues. The remaining two V regions differ from V_κ regions of the three subgroups and from each other^{24,25}. This survey suggests that the $V_\kappa 21$ group in BALB/c mice is coded for by at

*Present address: Biochemistry Department, Stanford Medical School, Stanford, California 94305.

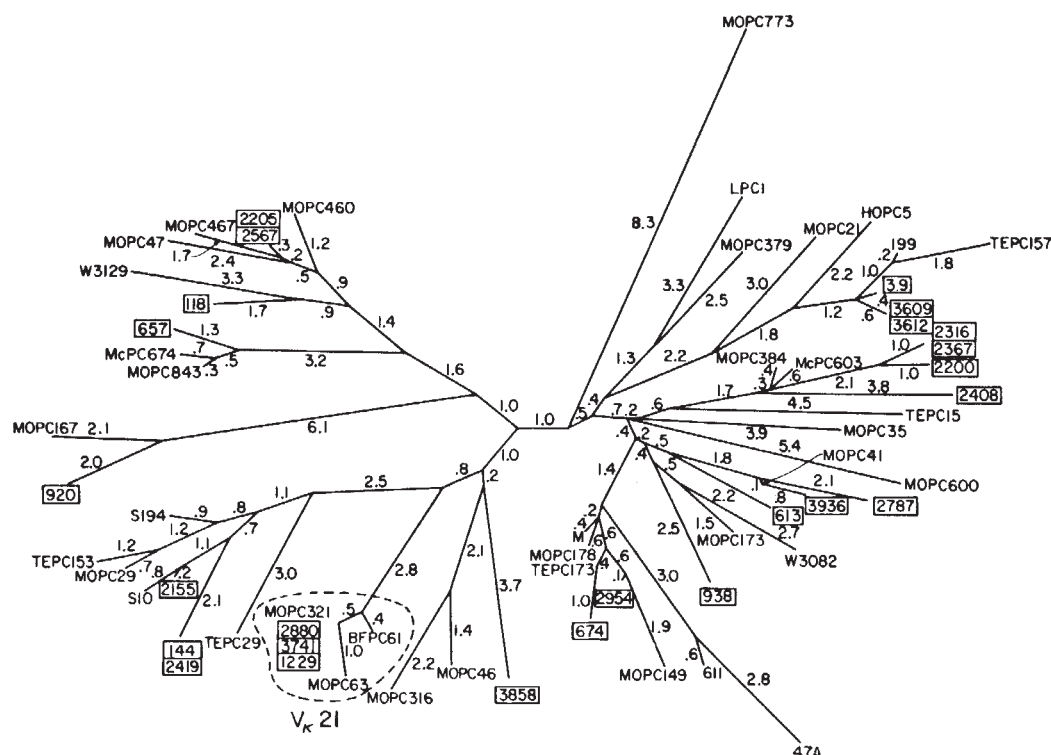


Fig. 1 A relatedness tree for N-terminal 23 residues of 55 different V_{κ} sequences of the mouse. These V_{κ} regions are derived from NZB (boxed) and BALB/c mice. The $V_{\kappa}21$ group is indicated by a dotted circle. The numbers on the lines indicate the number of nucleotide substitutions that separate two successive junction (node) points on the tree. Thus, the minimal extent to which individual κ chains differ from one another at the nucleotide level can be determined by summing the lengths of the lines between the two V_{κ} regions (adapted from ref. 19)

least three germline V genes by the argument of parallel mutation.

The object of this study is to analyse further the extent and pattern of diversity in the $V_{\kappa}21$ group of NZB mice. In a screen of the N-terminal sequences of 31 NZB κ chains, three $V_{\kappa}21$ chains, PC3741, PC2880 and PC1229, were found²⁶. Using the NZB κ chains, PC3741 and PC2880, or the BALB/c $V_{\kappa}21$ chains, T124 and M70, antisera were made which specifically identify $V_{\kappa}21$ chains and distinguish a set reacting with antisera to PC2880 or M70 from a second set reacting with antisera to PC3741 or T124 (Fig. 2). Of 700 NZB myeloma immunoglobulins screened with these antisera, 5.7% fell into the first set and 3.1% into the second. In addition, the normal light chains from the sera of several inbred strains of mice crossreact with these antisera to levels of 7.2% and 1.4%, respectively.²⁷ Thus, the $V_{\kappa}21$ group constitutes about 8–9% of normal serum and myeloma light chains. This group is an ideal model system for studying the nature of diversity within the κ family of the mouse because many members of the $V_{\kappa}21$ group can be readily identified by rapid serological assays. Furthermore, the high frequency of these κ chains in normal sera suggests that the $V_{\kappa}21$ light chains are used not only in myeloma antibodies, but also in normal antibodies. In this article we report on the V-region sequences of 22 NZB κ chains belonging to the $V_{\kappa}21$ group. These data suggest that the $V_{\kappa}21$ group is encoded by multiple germline genes and that the V region is encoded by two separate DNA segments, V and J.

$V_{\kappa}21$ regions fall into six subgroups

Twenty-two NZB $V_{\kappa}21$ regions are compared. A summary of the methodology used to sequence these V regions is given in the legend to Fig. 2 and the complete data will be presented in detail elsewhere^{28–30}. Of these $V_{\kappa}21$ regions, 18 fall into six subgroups, denoted by $V_{\kappa}21A$, B, C, D, E and F, in which two or more closely related V_{κ} regions share at least three subgroup-associated residues. Subgroup-associated residues distinguish the members of a particular subgroup from members of most of the other subgroups (boxed residues in Fig. 2); 29 subgroup-associated residues occur in positions 1–98. Variable regions from different subgroups differ from one another by 3 to 20 residues over their N-terminal 98 residues (Fig. 3). The reason for

limiting our subgroup analysis to the N-terminal 98 residues will become apparent later. Again, even closely related subgroups (such as $V_{\kappa}21E$ and $V_{\kappa}21F$) seem to be encoded by separate germline V genes by the argument of parallel mutation, namely, that it is unlikely that somatic mutation could generate in two or more independent lymphocyte lineages the three identical substitutions that separate V_{κ} regions of these two subgroups. Thus, at least one distinct germline V gene seems to encode each of the six subgroups (Fig. 3).

The minimal number of six $V_{\kappa}21$ germline genes deduced from V region amino acid sequences is larger than the one to three $V_{\kappa}21$ deduced from nucleic acid hybridisation experiments³¹. However, saturation-hybridisation studies using DNA probes derived from several $V_{\kappa}21$ mRNAs suggest that four to six germline genes encode the $V_{\kappa}21$ regions (see accompanying paper³²). This number is consistent with the number of V genes we have identified so far by subgroup analysis.

It is possible that additional $V_{\kappa}21$ subgroups will be found. Four of the NZB $V_{\kappa}21$ regions (that is, PC2154, PC2413, PC7461 and PC2960) do not fall into any of the six subgroups and are indicated as distinct branches on the relatedness tree (Fig. 3). These V regions cannot yet be designated separate subgroups because a similar reasoning is needed to invoke the argument that separate germline genes for each subgroup avoid multiple identical somatic mutations (see above). If new $V_{\kappa}21$ sequences are found which pair with these examples, additional $V_{\kappa}21$ subgroups would be defined. In addition, currently defined subgroups may split into two or more new subgroups. For example, PC2485 was a member of the $V_{\kappa}21E$ subgroup until the sequence of PC4039 was determined. These two identical sequences then became subgroup $V_{\kappa}21F$, sharing three distinct subgroup-associated residues (Fig. 2).

If additional $V_{\kappa}21$ subgroups are defined, then two possibilities must be considered. The saturation-hybridisation experiments have underestimated the number of $V_{\kappa}21$ genes. Alternatively, a somatic mechanism may exist which permits a single germline V gene to generate two or more subgroups of $V_{\kappa}21$ regions.

Genetic rearrangement during differentiation

The $V_{\kappa}21$ regions seem to have two distinct segments which we have designated the V segment (residues 1 to 98 or 99) and the J

The N- and C-terminal boundaries of the V and J segments are defined as follows. The J segment seems to have started by position 101 because of the invariant association of a leucine at this position with an alanine and leucine at positions 105 and 111, respectively, in the J segment of six different V κ 21 regions (Fig. 2). This repeated invariant association of three residues suggests that positions 101–111 behave as a discrete genetic unit. The last subgroup-associated residue of the V segment is at position 98. Accordingly, the N-terminal boundary of the J segment may start somewhere between residues 99 and 101. Position 100 is an amino acid insertion in one κ chain (PC7132) and we have arbitrarily chosen the J segment to include this

Nucleic acid studies on mouse λ and κ genes also suggest that there are V and J segments. Two mouse V_λ genes isolated from embryonic DNA terminate their V genes at a position homologous to residue 101 (refs 2, 18). In addition, a J_λ segment consisting of approximately 36 nucleotides and separated from the V_λ segment by intervening DNA sequences has been

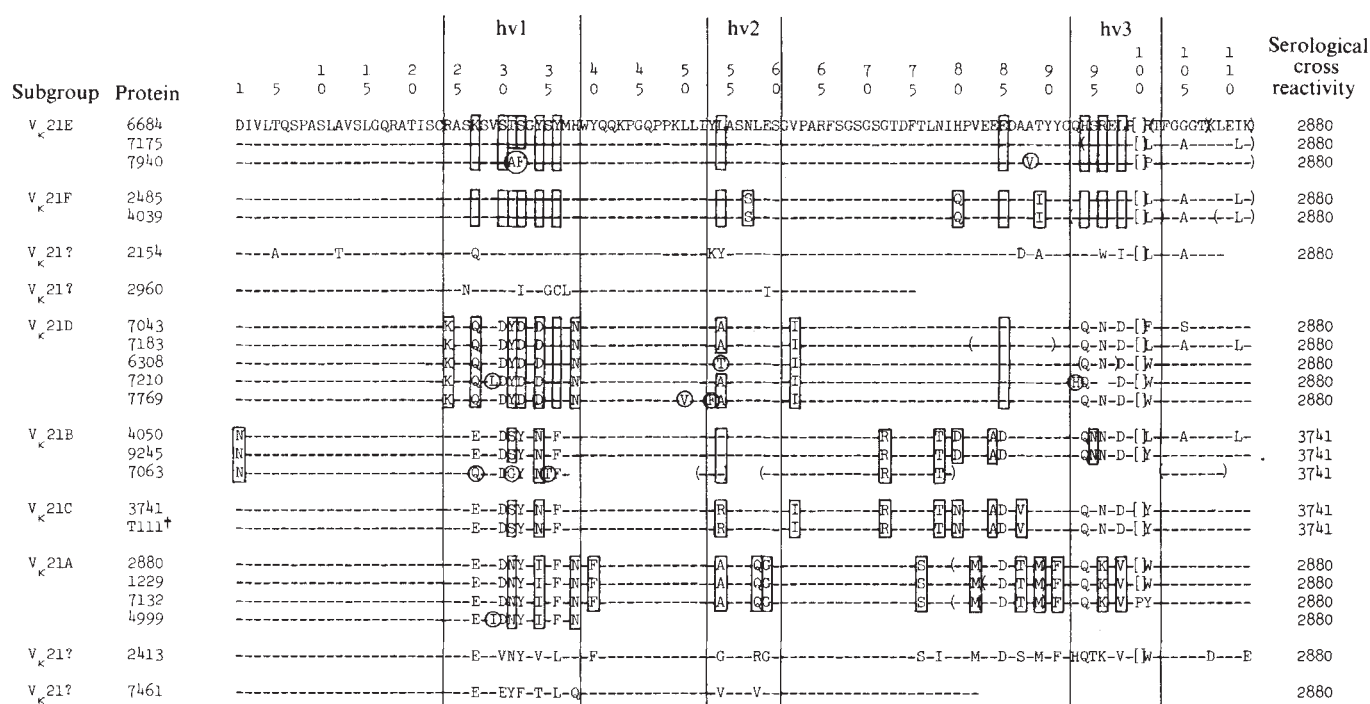


Fig. 2 The amino-acid sequences of V regions from the mouse V_κ21 group. The sequences are divided into subgroups which are designated V_κ21A–V_κ21F. Additional potential subgroups are indicated with question marks. The T111 sequence is a BALB/c protein²⁵, identical to PC3741, which permits the definition of subgroup V_κ21C. The subgroup-associated residues unique to three or fewer subgroups are boxed. The intra-subset substitutions are circled. Amino-acid compositional data are indicated by brackets. The three hypervariable regions are indicated as hv1, hv2 and hv3. The J segment extends along residues 100–112. The serological crossreactivity of individual proteins to antisera directed against two screening κ chains is given on the right. The following general strategy was used for sequencing these V regions. The intact light chain was analysed on the automatic sequenator for 40 residues. The procedures for automatic sequence analysis are given in ref. 28 where the data for a 78-residue run on PC4050 are presented. Methionine fragments were prepared; all V_κ regions but one have a methionine at position 37. Methionine fragments starting at position 38 were prepared for 40–65 residues. Only the V_κ21A subset has additional methionine residues and these fragments were isolated and sequenced. In certain cases mild acid cleavage was used to split aspartic acid–proline bonds. Arginine fragments extending over positions 65–113 were prepared for certain κ chains and sequenced for 35–40 residues. Thermolysin, tryptic and chymotryptic peptides were isolated from certain of these methionine or arginine fragments and compositions were determined. In some cases these peptides were sequenced in the presence of polybrene. A detailed analysis of these data will be presented elsewhere²⁸.

The V and J segments associate independently with one another (Fig. 4). In the 18 $V_{\kappa}21$ regions with relatively complete sequence data between residues 100 and 112, there are 16 different V segments and eight different J segments (Fig. 2). A particular J segment may be associated with V_{κ} segments from four different subgroups (sequence 5 in Fig. 4). Conversely,

The most attractive function for the J segment is that it may be involved in generating antibody diversity. Residues 100 and 101 constitute a portion of the third hypervariable region¹⁴. Accordingly, if the J segment occurs at positions 100–112, a

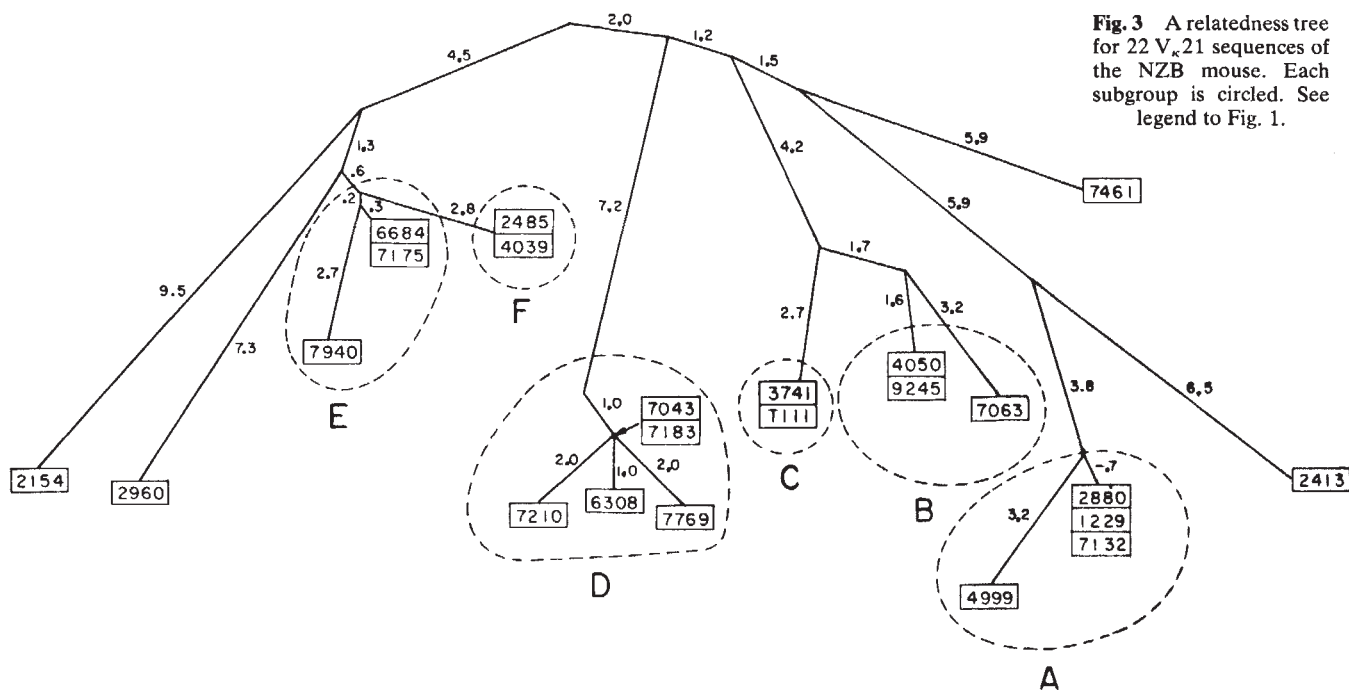


Fig. 3 A relatedness tree for 22 $V_{\kappa}21$ sequences of the NZB mouse. Each subgroup is circled. See legend to Fig. 1.

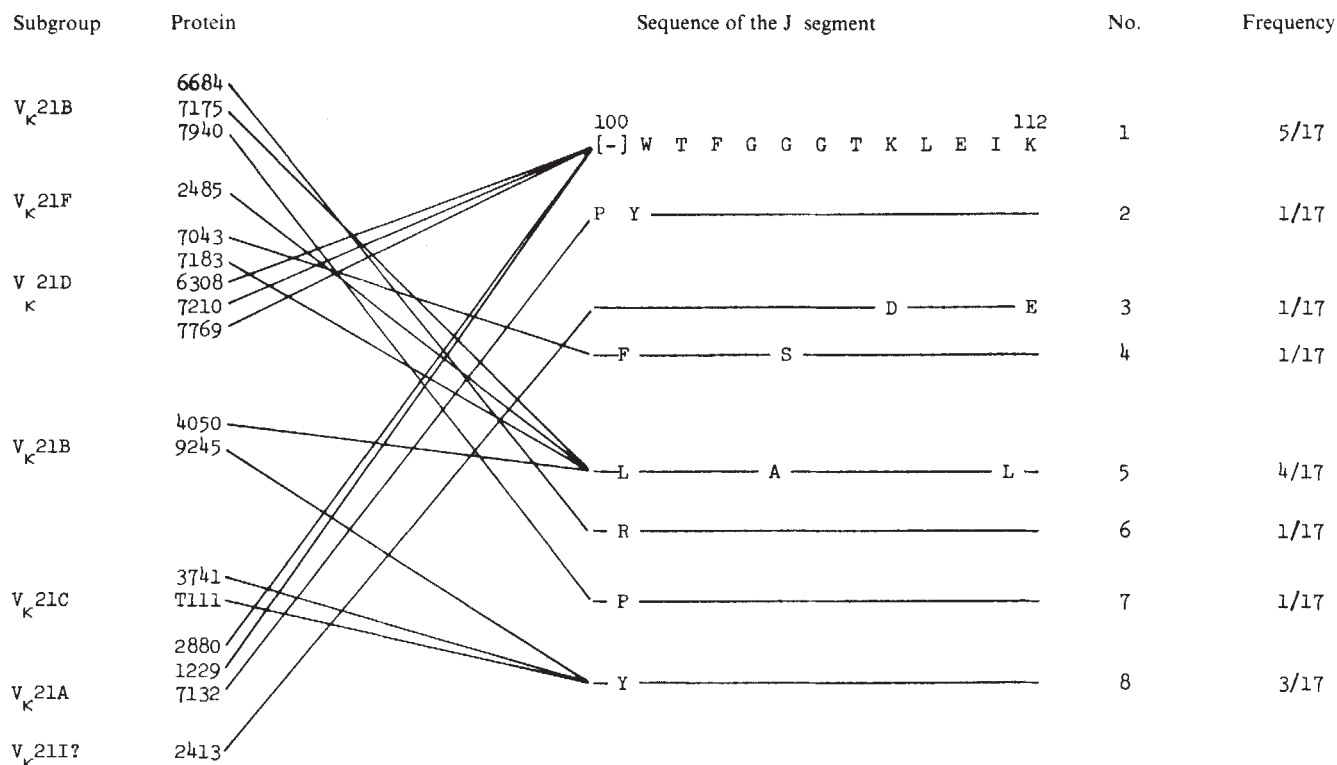
single J segment may be joined in a combinatorial fashion to each V segment to provide another stage in the amplification of antibody diversity. We describe this process as combinatorial joining. Indeed, position 101 is among the most variable in the $V_{\kappa}21$ regions examined (Fig. 2). This diversity might include sequence gaps (insertions or deletions) as well as nucleotide substitutions. The sequence gaps would result from J regions of different length (for example, proline at position 100 in PC7132).

The precise size of the J segment of $V_{\kappa}21$ genes must wait the DNA sequence analysis of $V_{\kappa}21$ genes from differentiated and

undifferentiated DNA. If J does not include part of the third hypervariable region, then combinatorial joining, as such, would not generate diversity in the combining site. In this case, invariant associations such as the leucine at 101 with alanine at 105 and leucine at 111 as well as the pattern of substitutions at position 101 require an alternative explanation.

The J segment may also mediate the joining of V, J and C gene information. DNA rearrangements and mRNA splicing³⁶⁻³⁸ must occur before the V, J and C gene information is joined into a single cytoplasmic mRNA (Fig. 5). The joining of V and J regions seems to occur at the DNA level¹⁸. The intervening

Fig. 4 A diagram illustrating the independent association of the V segments (residues 1-99) and the J segments (residues 100-112).



DNA sequence between the J and C regions could either lack other J regions (Fig. 5a) or contain them, in which case they would be eliminated by RNA processing (Fig. 5b).

Intra-subgroup variability

The analysis of the intra-subgroup diversity in mouse λ chains presented a simple mutational pattern consistent with somatic mutation¹⁵. Likewise, if each of the six $V_{\kappa}21$ subgroups is encoded by a single germline V gene, the nature and patterns of substitutions within each subgroup should reflect the mechanism of somatic mutation. The 12 intra-subgroup substitutions within the $V_{\kappa}21$ regions are indicated by circles in Fig. 2. Several features distinguish these V_{κ} substitutions from their V_{λ} counterparts discussed earlier. (1) Two of 12 substitutions are found outside the hypervariable regions. (2) Four of six variant sequences have two or more substitutions. If these differences arise by somatic mutation, the mutational mechanism operates outside as well as within hypervariable segments and shows a tendency to generate multiple substitutions in individual variant sequences. Moreover, somatic mutation followed presumably

sequences of myeloma proteins which may represent only a subset of the total V genes. Our studies suggest the $V_{\kappa}21$ group is encoded by at least six V genes. In addition, studies at the DNA level on three V_{κ} groups suggest that four to eight germline V genes are present in each group^{23,32}. If we assume that seven germline V genes are present in each group defined by N-terminal analysis, then approximately 350 germline genes encode the V_{κ} family. Less is known about the diversity in the V_H gene family, but a similar estimate of at least 200 V_H genes seems reasonable.³⁹ In contrast, the mouse λ gene family is apparently encoded by one or two germline genes.

Combinatorial association: That most light chains may combine with most heavy chains is suggested by several types of observations^{19,40,41}. This non-genetic mechanism for generating antibody diversity is termed combinatorial association. In principle, 350 V_L and 200 V_H regions could, through combinatorial association, generate 7×10^4 different types of antibody molecules. If the mammalian antibody repertoire is greater than 10^6 (see ref. 42), and if we assume that there are only 350 V_L and 200 V_H germline genes, then more than 90% of the antibody repertoire must be generated by somatic mechanisms.

Somatic mutation: All the intra-subgroup substitutions noted

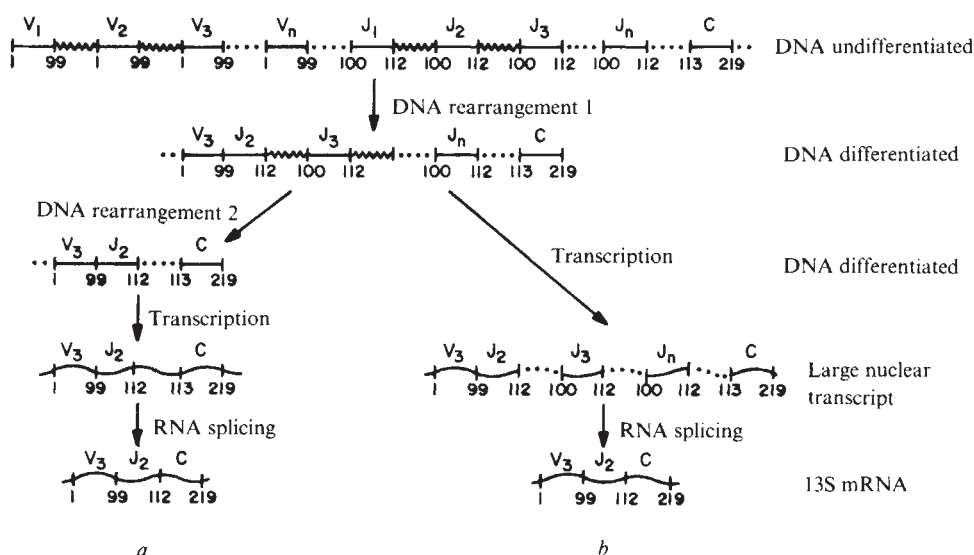


Fig. 5 Hypothetical models for the DNA segments of mouse κ genes. These gene segments are rearranged at the DNA level, presumably as the lymphocyte differentiates. As with λ genes, DNA rearrangement is depicted as joining the V and J segments⁴. There are at least two pathways for joining the J and C segments at the nucleotide level. *a*, A second DNA rearrangement moves the J_2 segment closer to the C segment, possibly deleting the intervening J sequences. A recent report suggests that two or more DNA rearrangement events can occur in lymphocytes during the switch in synthesis from one to a different immunoglobulin class⁴⁸. Final joining occurs by mRNA splicing. *b*, A large nuclear RNA transcript complete with the intervening J segments is processed by mRNA splicing enzymes.

by antigen selection seems to favour diversity at certain sites in the hypervariable regions.

Again, this analysis of intra-subgroup variability must be qualified by the possibility that one variant differing by two or more residues from the prototype sequence will become a distinct subgroup when a second identical sequence is determined. Thus, perhaps the reason the intra-subgroup variability in the $V_{\kappa}21$ chains and the λ chains differ is that in the former some of the variants are encoded by additional germline V genes.

Sources of antibody diversity

The analysis of the $V_{\kappa}21$ group of mouse V_{κ} regions offers certain insights into each of four basic mechanisms for producing antibody diversity.

Multiple germline V genes: According to a statistical analysis of N-terminal data, there are about 50 κ groups^{15,39}. This number may be an underestimate because it is based on the

in the $V_{\kappa}21$ regions are single-base substitutions, predominantly in the hypervariable regions. Obviously, some or all of this variation could arise by spontaneous somatic mutation or by a hypermutational mechanism.

Combinatorial joining: One potential source of somatic diversification is the combinatorial joining of multiple J segments and V segments (Fig. 5). Thirteen different J segments have already been identified in mouse κ chains. For example, if 10 different J segments may be joined to any of 350 different V segments, then 3,500 different V regions may be generated. If combinatorial joining of a similar magnitude occurs in light and heavy variable regions, 7×10^6 antibody molecules could be generated by this process.

Combinatorial joining could generate identical variants among mice of the same inbred strain. In contrast, most somatic theories have assumed that the generation of diversity is a random process occurring by mutation, recombination or gene conversion. A reproducible somatic process could explain the

patterns of diversity seen in certain immune responses. For example, the responses to $\alpha 1,3$ -glucosyl linkages of BALB/c mice⁴³ and to the 4-hydroxy-3-nitrophenylacetyl hapten (NP) of C57BL/6 mice⁴⁴ are heterogeneous by isoelectric focusing, but the heterogeneity is limited to V_H and is reproducible in each case^{45,46}. Perhaps this reproducible heterogeneity arises from the combinatorial joining of V_H -DEX or V_H -NP^b genes with a variety of different J_H segments. A similar possibility could explain the closely related heavy-chain sequences produced in response to the immunisation of guinea pigs against several haptens⁴⁷.

Other investigators have proposed combinatorial joining as a mechanism for antibody variability^{7,8}, but their models arose from the conviction that certain hypervariable regions may be associated with many different framework regions. For example, based on a computer analysis of all available V-region sequences, Kabat *et al.* have suggested the mini-gene model, which proposes that each of the three hypervariable regions and the four intervening framework regions are encoded by mini-genes which are joined during differentiation⁹. The correlation of specific hypervariable and framework segments within certain V_{κ} 21 subgroups (Fig. 2) argue against this general model. Our data suggest that Kabat's model is correct in principle, but that the combinatorial process occurs only within the third hypervariable region. The identification of a J segment in λ chains raises the possibility that combinatorial joining of V and J segments will be used by all three antibody families, λ , κ and H.

Conclusion

Antibody variability may be generated by a variety of mechanisms, some genetic (multiple V genes, combinatorial joining and somatic mutation) and others non-genetic (combinatorial association). Combinatorial joining is both a somatic and a germline mechanism in that it presumably rearranges germline information during somatic differentiation. The critical question that remains for the future is what are the relative contributions of these four basic mechanisms for generating diversity to the functional repertoire of antibody molecules? The answer to this intriguing question will obviously require combined chemical, biological and genetical approaches.

This research was supported by a grant from the NSF grants from the NIH and an appropriation from the Commonwealth of Pennsylvania. We thank Drs Michael Potter and David McKean who developed the serological methods essential for identifying V_{κ} 21 light chains.

Received 22 August; accepted 31 October 1978.

1. Tonegawa, S., Hozumi, N., Mathysens, G. & Schuller, R. *Cold Spring Harb. Symp. quant. Biol.* **41**, 877-889 (1976).
2. Tonegawa, S., Maxam, A. M., Tizard, R., Bernard, O. & Gilbert, W. *Proc. natn. Acad. Sci. U.S.A.* **75**, 1485-1489 (1978).
3. Hozumi, N. & Tonegawa, S. *Proc. natn. Acad. Sci. U.S.A.* **73**, 3628-3632 (1976).
4. Brach, C., Hiram, M., Lenhard-Schuller, R. & Tonegawa, S. *Cell* (in the press).
5. Hood, L., Campbell, J. H. & Elgin, S. C. R. *A. Rev. Genet.* **9**, 305-353 (1975).
6. Weigert, M., Cesari, I. M., Yonkovich, S. J. & Cohn, M. *Nature* **228**, 1045-1047 (1970).
7. Wu, T. & Kabat, E. *J. exp. Med.* **132**, 211-250 (1970).
8. Capra, J. D. & Kindt, T. *J. Immunogenetics* **1**, 417-427 (1975).
9. Kabat, E. A., Wu, T. T. & Bilofsky, H. *Proc. natn. Acad. Sci. U.S.A.* **75**, 2429-2433 (1978).
10. Potter, M. *Physiol. Rev.* **52**, 631-719 (1972).
11. Kabat, E. A., Wu, T. T. & Bilofsky, H. *Variable Regions of Immunoglobulin Chains, Tabulations and Analyses of Amino Acid Sequences* (Bolt, Bearanek and Newman, Massachusetts, 1976).
12. Amzel, L., Polyjak, R., Saul, F., Varga, J. & Richard, F. *Proc. natn. Acad. Sci. U.S.A.* **71**, 1427-1430 (1974).
13. Segal, D. M. *et al. Proc. natn. Acad. Sci. U.S.A.* **71**, 4298-4302 (1974).
14. Dugan, E. S., Bradshaw, R. A., Simms, E. S. & Eisen, H. N. *Biochemistry* **12**, 5400-5416 (1973).
15. Weigert, M. & Riblet, R. *Cold Spring Harb. Symp. quant. Biol.* **41**, 837-846 (1976).
16. Tonegawa, S. *Proc. natn. Acad. Sci. U.S.A.* **73**, 203-207 (1976).
17. Leder, P., Honjo, T., Swan, D., Peckman, S., Nau, M. & Norman, B. in *Molecular Approaches to Immunology, Miami Winter Symposia*, Vol. 9 (eds Smith, E. E. & Ribbons, D. W.) 173-188 (Academic, New York, 1975).
18. Bernard, O., Hosumi, N. & Tonegawa, S. (submitted).
19. Hood, L. *et al. Cold Spring Harb. Symp. quant. Biol.* **41**, 817-836 (1976).
20. Potter, M. *Adv. Immun.* **25**, 141-212 (1978).
21. Hood, L., Gray, W., Sanders, E. & Dreyer W. *Cold. Spring Harb. Symp. quant. Biol.* **32**, 133-146 (1967).
22. Milstein, C. *Nature* **216**, 330-332 (1967).
23. Rabbitts, T. H. *Immun. Rev.* **36**, 29-50 (1977).
24. McKean, D., Potter, M. & Hood, L. *Biochemistry* **12**, 749-759, 760-771 (1973).
25. McKean, D. J., Bell, M. & Potter, M. *Proc. natn. Acad. Sci. U.S.A.* (in the press).
26. Loh, E. *et al. Proc. natn. Acad. Sci. U.S.A.* (in the press).
27. Julius, M., Potter, J., McKean, D. & Weigert, M. (submitted).
28. Hunkapiller, M. & Hood, L. *Biochemistry* **17**, 2124-2133 (1978).
29. Loh, E., Schilling, J., Weigert, M. & Hood, L. (submitted).
30. Schilling, J., Loh, E., Weigert, M. & Hood, L. (submitted).
31. Tonegawa, S., Steinberg, C., Dube, S. & Bernardini, A. *Proc. natn. Acad. Sci. U.S.A.* **71**, 4027-4031 (1974).
32. Valbuena, O., Marcu, K. B., Weigert, M. & Perry, R. P. *Nature*, **276**, 780-784 (1978).
33. Leder, P., Honjo, T., Seidman, J. & Swan, D. *Cold Spring Harb. Symp. quant. Biol.* **41**, 855-862 (1976).
34. Rabbitts, T. & Milstein, C. *Cont. Topics molec. Immun.* **6**, 117-143 (1977).
35. Seidman, J. G., Leder, A., Nau, M., Norman, B. & Leder, P. *Science* (in the press).
36. Gilmore-Herbert, M. & Wall, R. *Proc. natn. Acad. Sci. U.S.A.* **75**, 342-345 (1978).
37. Schibler, U., Marcu, K. & Perry, P. (submitted).
38. Rabbitts, T. H. & Forster, A. *Cell* **13**, 319-327 (1978).
39. Cohn M. *et al. in The Immune System, Genes, Receptors, Signals* (eds Sercarz, E. E., Williamson, A. R. & Fox, C. F. 89-118 (Academic, New York, 1974).
40. Poljak, R. J., Amzel, L. M., Chen, B. L., Phizackerly, R. P. & Saul, F. *Immunogenetics* **2**, 393-394 (1975).
41. Milstein, C. *et al. Cold Spring Harb. Symp. Quant. Biol.* **41**, 793-803 (1976).
42. Williamson, A. R. *A. Rev. Biochem.* **45**, 467-500 (1976).
43. Blomberg, B., Geckeler, W. & Weigert, M. *Science* **177**, 178-180 (1972).
44. Imanishi, T. & Mäkelä, O. *J. Exp. Med.* **140**, 1498-1510 (1974).
45. McMichael, A. J., Phillips, J. M., Williamson, A. R., Imanishi, T. & Mäkelä, O., *Immunogenetics* **2**, 161-173 (1975).
46. Hansburg, D., Briles, D. E. & Davie, J. M. *J. Immun.* **119**, 1406-1412 (1977).
47. Cebra, J. J., Koo, T. H. & Ray, A. *Science* **186**, 263-265 (1974).
48. Honjo, T. & Kataoka, T. *Proc. natn. Acad. Sci. U.S.A.* **75**, 2140-2144 (1978).

The arrangement and rearrangement of antibody genes

J. G. Seidman & Philip Leder

Laboratory of Molecular Genetics, National Institute of Child Health and Human Development, National Institutes of Health, Bethesda, Maryland 20014

Cloned segments of mouse chromosomal DNA provide direct evidence for the somatic rearrangement of κ variable and constant region genes in antibody producing cells. This rearrangement apparently affects only one member of an allelic pair of light chain genes.

THE notion that immunoglobulin genes are encoded in several segments that rearrange during somatic differentiation has been incorporated into most theories of antibody diversity since the 'two gene-one polypeptide' argument was first advanced by Dryer and Bennet in 1965 (ref. 1). Most recently, certain details of this arrangement have become clear through a series of experiments involving the analysis of fragments containing

chromosomal DNA encoding variable and constant region immunoglobulin genes. As predicted, constant and variable regions are separately encoded in germline DNA²⁻⁴ and appear to be brought together by a somatic recombination event⁵⁻⁷. Moreover, it seems likely that this recombination is closely and critically related to light chain production; that is, that the rearranged variable region sequence will be the one expressed in a given immunocyte.

Further analysis of these genes, several of which have been cloned, indicates that their organisation is even more complex than originally suspected. The inactive κ and λ mouse variable region genes sequenced thus far fail to encode the last 10 or so amino acids that had, by sequence analysis, always been associated with the variable region^{4,8}. In a direct examination of a

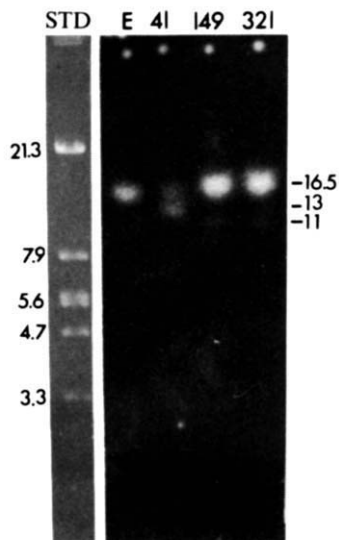


Fig. 1 Analysis of the *Bam*H1 fragments of BALB/c embryonic and plasmacytoma DNAs annealed to a constant region gene sequence. The fragments obtained by *Bam*H1 digestion of embryonic (E), MOPC-41 (41), MOPC-149 (149) and MOPC-321 (321) DNA were fractionated by electrophoresis (from top to bottom) in a 0.7% agarose gel, denatured, neutralised, transferred to nitrocellulose sheets and hybridised with a 32 P-labelled DNA fragment (136 c.p.m. per pg) which contained constant region sequences as described previously⁴. The constant region containing DNA fragment used as probe was the 2.8-kilobase *Bam*H1–*Hind*III fragment indicated in Fig. 6, which contains the κ constant region gene. The embryo specific band (16.5 kilobase) and the plasmacytoma specific bands (13.0 and 11.0 kilobases) are indicated. Fragment lengths were determined by comparison with ethidium bromide-stained, *Eco*R1 fragments of phage λ DNA of indicated length (STD).

cloned embryonic mouse λ light chain DNA fragment, Tonegawa and his associates have shown that this terminal variable region segment, called the J sequence, is separately encoded⁹. While the precise borders of these regions have not yet been determined, it seems that some—if not all—light chain genes are encoded in at least three segments, a major variable region sequence, a short J sequence and the constant region sequence. This observation also seems to fit in with recent comparisons of amino acid sequence data (ref. 10 and M. G. Weigert and L. Hood, personal communication).

Since the somatic rearrangement of these three segments may play a part in generating antibody diversity as well as in antibody gene activation, it will be important to determine whether gene rearrangement is a constant feature of immunocyte development and to understand its precise molecular basis. In some respects this is most advantageously investigated in the mouse κ light chain system where hundreds of variable region genes exist in the embryonic chromosome^{4,11} and are presumably candidates for rearrangement and, hence, expression.

In what follows we describe the cloning and characteristics of several immunoglobulin κ light chain genes derived from the chromosomes of embryonic and κ chain producing cells. While different gene arrangements appear in each producing myeloma cell line, the basic embryonic gene arrangement is also preserved in each. One of the rearranged fragments contains an entire light chain sequence in which variable and J sequences appear contiguous and are separated from the constant region gene by a 3.7-kilobase intervening sequence of DNA. Several cloned κ gene fragments have also recently been obtained by Lennard-Schuller *et al.*²¹.

Cloning fragments containing constant region genes

By using hybridisation probes derived from cloned κ light chain cDNA, we were previously able to identify several *Eco*R1 fragments of embryonic and myeloma DNA (MOPC-149 and

MOPC-41) that encoded closely related variable region gene sequences¹¹. In contrast, each tissue yielded what appeared to be a single *Eco*R1 fragment of DNA encoding a constant region sequence that curiously migrated at approximately the same position in a two-dimensional analysis involving RPC-5 chromatography and gel electrophoresis¹¹. Evidently, whatever rearrangement occurred amongst these genes, it resulted in the production of new DNA sequences that closely resembled the embryonic *Eco*R1 pattern in length and chromatographic properties.

That the gene rearrangement had actually occurred in MOPC-149 and MOPC-41 cells could be demonstrated using a different restriction enzyme to produce a new fragment pattern in the producing cells. DNA from embryonic, MOPC-41, MOPC-149 and MOPC-321 cells were digested with *Bam*H1, electrophoresed, blotted to nitrocellulose sheets¹² and annealed to a radioactive probe corresponding to the κ constant region (Fig. 1). Each tissue, including embryo, yielded a 16.5-kilobase DNA fragment that contained constant region sequences. Each myeloma, on the other hand, yielded an additional fragment, in the case of MOPC-41, approximately 13 kilobases long, in the case of MOPC-149 and MOPC-321, 11 kilobases long. (MOPC-321 was previously analysed with *Bam*H1 by Hozumi and Tonegawa⁵ and was included as a control. The constant fragments we observe differ in size from those seen by the earlier workers and retain the germline fragment which was apparently not observed in their analysis. It is possible that we are dealing with different tumours.) Given these results, it seemed likely

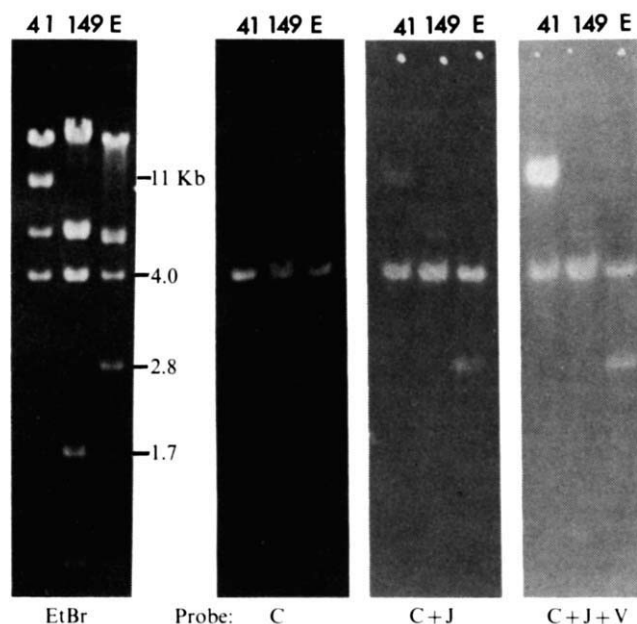


Fig. 2 Restriction endonuclease *Hind*III digestion of cloned embryonic and plasmacytoma DNA fragments encoding κ gene sequences. The fragments obtained by *Hind*III digestion of λ Ch4A phage containing 41KC-21, 149C-2 and EC-5 *Eco*R1 DNA fragments were fractionated on a 1.2% agarose gel and stained with ethidium bromide (EtBr). The *Hind*III fragments (left) were denatured, transferred to nitrocellulose sheets and hybridised separately to three 32 P-labelled probes and autoradiographed (right) as described previously^{4,11}. The probes were 32 P-labelled fragments containing segments of the κ light chain. The constant region probe was a *Hpa*I/*Bam*H1 fragment obtained from pBR322-K41 (ref. 11) which should encode from amino acids 125 through the 3' untranslated end. The J region (J+C) probe was a *Hinc*II fragment obtained from pCR1-K149 (ref. 19) which encodes amino acids 90–125 of the MOPC-149 light chain. The variable region probe (V+J+C) was a *Bam*H1/*Hinc*II fragment cleaved out of pBR322-K41 (ref. 11) which encodes amino acids 44–125 of the MOPC-41 light chain. The sizes in kilobases of the *Hind*III fragments were determined relative to phage λ *Eco*R1 fragments. 149KC-2 does not have a separate fragment that hybridises to the J region probe, while 41KC-21 has both a J region and a variable region gene sequence.

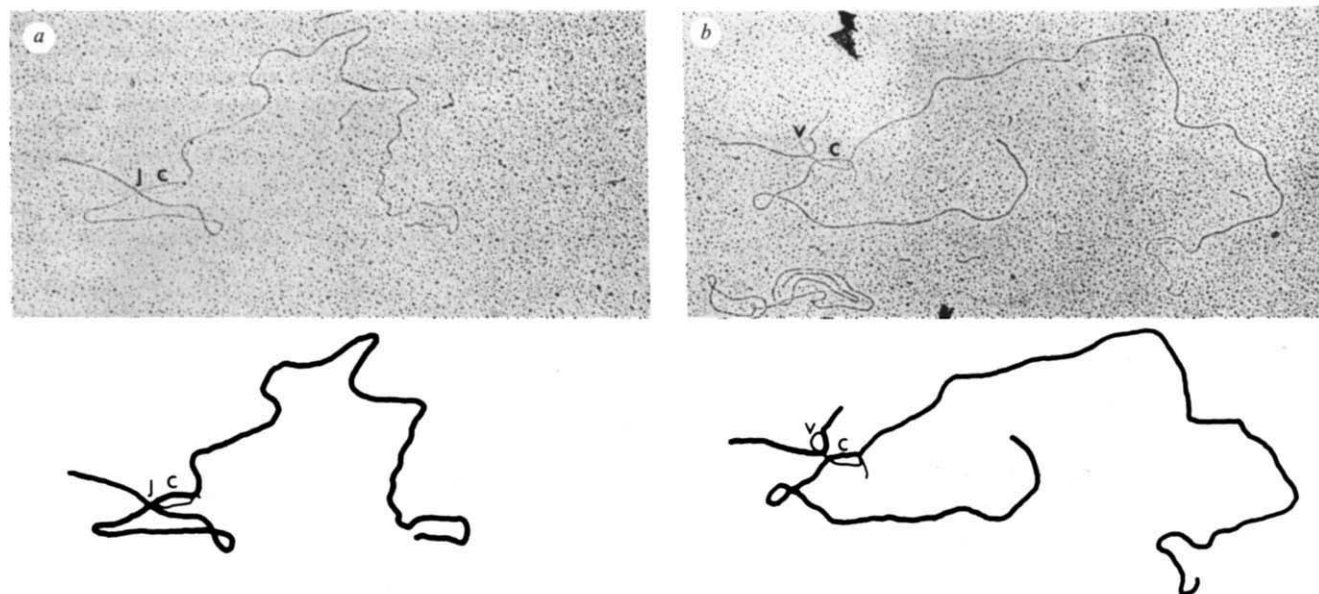


Fig. 3 Visualisation of the κ light chain sequences encoded by two cloned *EcoRI* fragments derived from embryonic DNA and MOPC-149 DNA. *a*, The 16.0-kilobase *EcoRI* fragment isolated from the phage λ Ch4A-EC-5 was annealed to MOPC-41 mRNA and spread for electron microscopic analysis as described previously²¹. *b*, The 14.9-kilobase *EcoRI* fragment ($20 \mu\text{g ml}^{-1}$) from the phage λ Ch4A-149C-2 and an 1800-base pair *HaeIII* fragment ($20 \mu\text{g ml}^{-1}$), from pMB9-K2 which encodes a MOPC-149-like variable region⁴ were annealed with a MOPC-149 mRNA ($10 \mu\text{g ml}^{-1}$) in 70% formimide, 0.5 M NaCl, 0.1 M Tricine pH 8.0, 0.01 M EDTA at 52°C for 18 h, and spread for electron microscopy. The R loops are interpreted by diagrams drawn immediately below each molecule. The constant region R loop (C), 475 ± 78 base pairs long and the variable region R loop (V) is 299 ± 68 base pairs long. The analysis of 35 molecules suggested that the constant region R-loop is 10.4 ± 0.24 kilobases, from one end and 5.1 ± 0.20 kilobases from the other end of the fragment. The location of the J region gene was determined by the distance, 1.3 ± 0.19 kilobases, from the end of the fragment to the point where the DNA falls across the end of the R loop. The long ends of both DNA fragments are single stranded because of nicks introduced into the DNA by nuclease in the *EcoRI* preparation. Although some of the *EcoRI* fragments had single stranded ends nearly all of the fragments were full length when the single strand DNA was included in the length of the fragment.

that the *EcoRI* digestion of the DNA from each tumour produced a mixture of germline and rearranged fragments of approximately the same size. Obviously their structural features could best be resolved by cloning.

The *EcoRI* fragments from embryonic and plasmacytoma DNA which contained constant region gene sequences were therefore cloned, using procedures previously used to clone the mouse globin¹³⁻¹⁵ and κ variable region genes⁴. *EcoRI* fragments containing κ constant region sequences were purified first by RPC-5 column chromatography and then by preparative gel electrophoresis, ligated to the DNA of the EK2 vector λ Ch4A (ref. 16) and finally packaged into phage particles using the procedure developed by Becker and Gould¹⁷ and Sternberg, Tiemeier and Enquist¹⁸ as modified by F. Blattner and his co-workers (personal communication). In this way, 8×10^4 hybrid phage were recovered carrying embryonic DNA fragments. Of these, 20 contained constant region sequences. Twenty thousand hybrid phage were derived carrying MOPC-149 DNA fragments. Four contained constant region sequences. Ten thousand hybrid phage were derived carrying MOPC-41 DNA fragments. Three contained constant region sequences.

Light chain producing cells contain both embryonic and rearranged constant region genes: restriction enzyme mapping

Antibody production seems to involve the expression of only a single allele of the pair of genes presumably present in diploid cells. While various potential mechanisms might be invoked to account for this allelic exclusion, it is appealing to consider that the rearrangement of only one member of an allelic pair might create only one complete and functional copy of light and heavy chain genes. We have already noted that the constant region *BamHI* restriction fragment pattern derived from the plasmacytoma cells reveals one fragment that is the same size as that seen in embryonic DNA (Fig. 1) If only one of the constant

region alleles has undergone rearrangement, we might expect to find, in addition to the rearranged fragment, several fragments among our plasmacytoma-derived clones that are identical to those cloned from embryonic DNA, that is, non-rearranged fragments. To test this possibility, the DNA from five of the embryo-derived constant region gene containing phage and DNA from each of the phage carrying MOPC-41 and MOPC-149 constant region sequences were cleaved with several restriction enzymes and the resulting fragments were fractionated on agarose gels (data not shown). Not surprisingly, all five phage DNAs isolated from embryonic DNA were apparently identical (except with respect to their orientation within the vector). Furthermore, three of the four MOPC-149 derived phage and two of the three MOPC-41 derived phage seemed to be identical to the phage derived from embryonic DNA. It is possible that some of these embryo-like fragments were derived from non-plasmacytoma cells present in tumour tissue. This seems less likely in view of the minor degree of tissue contamination observed microscopically and the prevalence of embryo-like fragments among hybrid clones.

Although several of the cloned fragments obtained from plasmacytoma DNA were apparently identical to that derived from embryonic DNA, one cloned fragment from each light chain producing cell differed from the embryonic DNA fragment. These two isolates, derived from plasmacytoma MOPC-149 and MOPC-41 DNAs, were designated 149KC-2 and 41KC-21, respectively. (The phage isolate containing the embryo-derived *EcoRI* fragment which was used for comparison in these studies was called EC-5.) When the phage DNAs were cleaved with *EcoRI* and fractionated on a 0.6% agarose gel, small differences in the sizes of the three constant region containing fragments could be detected; the fragments EC-5, 149KC-2 and 41KC-21 were 16.0, 14.9 and 16.5 kilobases respectively (data not shown). While these differences became more obvious when the hybrid phage were cleaved with the restriction endonuclease *HindIII* (Fig. 2), the differences between the fragments were not great, suggesting that there was

substantial homology between them. Indeed, most enzyme digests produced several fragments that were identical in length among the three phage and yielded only one or two fragments that actually differed.

To determine which of the restriction fragments contain portions of the light chain structural gene, the separated fragments were denatured in the gel, transferred to nitrocellulose sheets and hybridised to the appropriate ^{32}P -labelled probes corresponding to different segments of the κ light chain mRNA. An example of results obtained in this way is shown in Fig. 2. A single 4.0-kilobase *Hind*III fragment anneals to a cloned constant region cDNA probe (extending from codon position 125 through the 3' untranslated sequence) and is found in all three genomic fragments. Surprisingly, a probe incorporating the variable-constant region junction (codon 90–125), presumably including the suspected J sequence, hybridised to two fragments in the digest of the embryonic clone, indicating that this sequence is interrupted by *Hind*III site in the embryonic genome. The second fragment is absent from MOPC-149 clone, but is present on the clone derived from MOPC-41 DNA (Fig. 2). Apparently, this particular J region gene sequence has been lost in the chromosomal rearrangement that produced the new MOPC-149 fragment.

Since two of these cloned *Eco*R1 fragments (from embryo and MOPC-41 DNA) contained both a J region gene and a constant region gene, either might also carry a variable region gene. The fragment derived from MOPC-149 DNA, while rearranged, failed to hybridise to MOPC-149 variable region coding sequence (data not shown). Therefore, as we had already concluded from previous analyses of genomic DNA⁴, an *Eco*R1 site separates variable and constant region genes in the MOPC-149 genome. On the other hand, the MOPC-41 V+J+C region gene probe hybridises much more strongly the 11.0-kilobase *Hind*III fragment derived from the MOPC-41 hybrid phage than did the probe containing only the J and constant region sequences. This suggests that the new MOPC-41 *Eco*R1 fragment might contain the variable region as well as the J and the constant region sequences (see below). That is, the entire MOPC-41 light chain structural gene might be encoded in this single *Eco*R1 fragment.

Visualisation of embryonic and rearranged MOPC-149 gene sequences

In situ hybridisation analysis indicated that the three *Eco*R1 constant region fragments contained extensive regions of homology, but were nevertheless different from one another. The embryonic clone—which was repeatedly discovered among cloned fragments derived from each plasmacytoma—encoded a constant region sequence and a J sequence homologous to that represented in MOPC-41 cDNA. The rearranged DNA segment derived from MOPC-149 DNA encoded a constant region sequence, but had evidently lost the nearby J region segment and, as a new *Eco*R1 site interceded between constant and variable region, a MOPC-149 variable region sequence did not occur on our cloned fragment. The rearranged DNA segment derived from the MOPC-41 DNA, on the other hand, seemed to encode all three segments. (These features are summarised in the map of each fragment represented in Fig. 6.)

The presence of each of these structural features could be confirmed and their relative position on each fragment determined by direct visualisation of R-loop structures formed by annealing light chain (MOPC-149 or MOPC-41) mRNA to each segment. In appropriate conditions, the light chain mRNA anneals to complementary sequences in one of the DNA strands forming a double-stranded RNA–DNA hybrid and displacing the other strand to form a bubble-like structure called an R loop²⁰. When the embryonic fragment (EC-5) was examined in this way, a single R loop 475 ± 78 base pairs long was observed, located 5.1 ± 0.20 and 10.4 ± 0.24 kilobases from the ends of the fragment (Fig. 3a). A similar structure was observed when the fragment derived from MOPC-149 was analysed (Fig. 3b). In

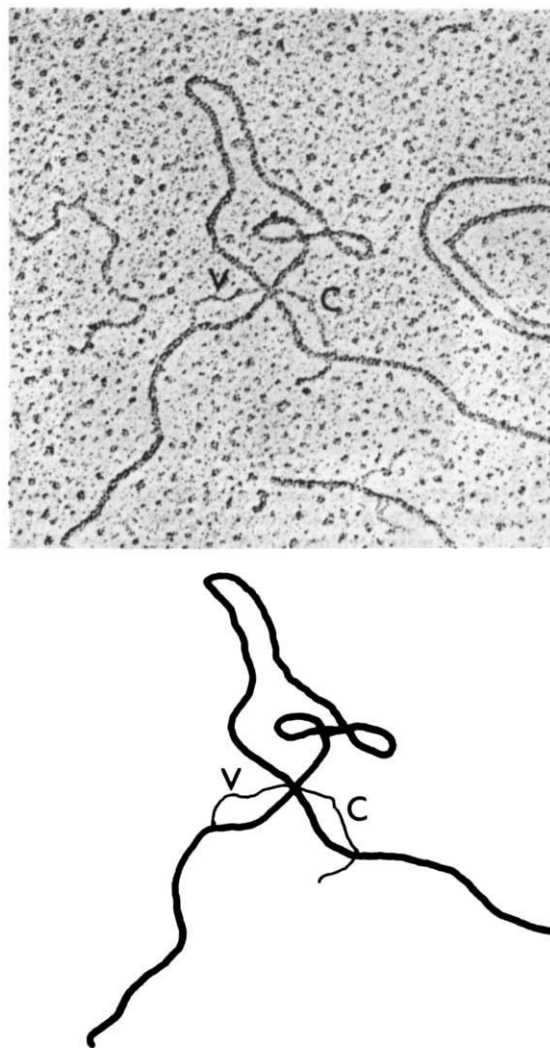


Fig. 4 The R loop formed between a fragment cloned from MOPC-41 DNA and MOPC-41 mRNA. The variable (V) and constant (C) region R loops formed as described previously⁴ are separated by 3.7 ± 0.187 kilobases of DNA. Twenty-five molecules were measured.

both cases, the loop structure should represent the constant region sequence (Fig. 2). This could be confirmed and the gene orientated (5'–3') by forming a second R loop using a previously cloned DNA fragment known to encode a MOPC-149-like variable region gene⁴ (Fig. 3b). Here the light chain mRNA links the two cloned fragments by forming an R loop on the smaller variable region fragment (1.8-kilobase *Hae*III fragment of pMB9-K2)⁴ and the larger constant region fragment. The smaller fragment serves as a marker to identify the 5' side of the constant region R loop and confirms that the constant region sequence is annealed to the larger, MOPC-149 fragment. Similar structures were visualised using the embryonic fragment.

The J region sequence on the embryonic fragment could also be identified by electron microscopic visualisation, despite its small size. The J region is too small to form a visible R loop, but could form a stable hybrid when annealed to MOPC-149 mRNA (Fig. 3a). Our hybridisation studies already suggested that this sequence was not immediately adjacent to the constant region sequence (Fig. 2). Indeed, if the J region were encoded at a distance from the constant region gene, we would expect to find molecules containing a constant region R loop and an intervening loop of double-stranded DNA joined to the constant region R loop by the J sequence in the mRNA. We identified 10 such structures in which a ~ 3.7 -kilobase length of

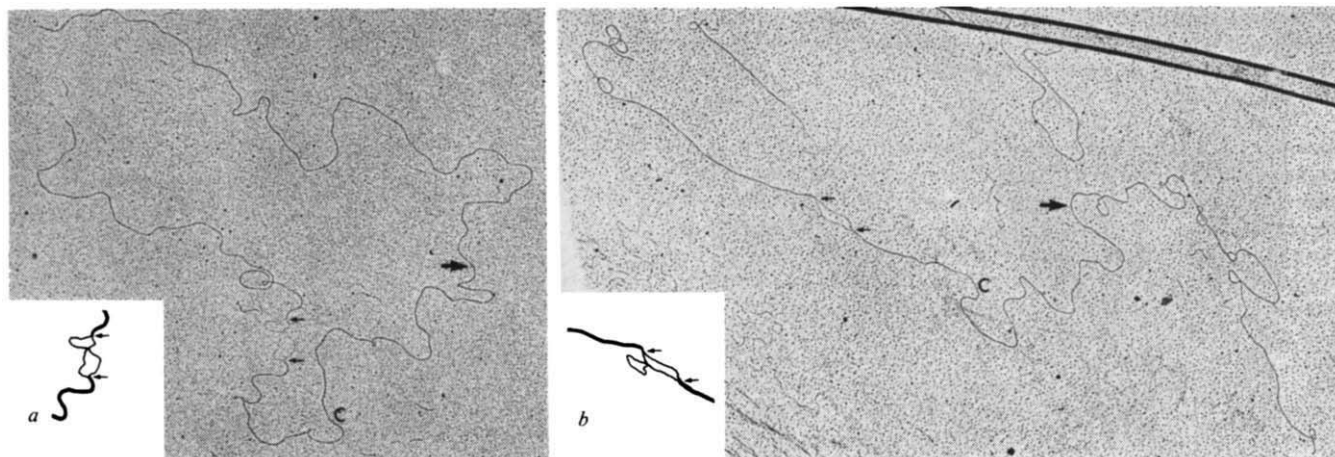


Fig. 5 Heteroduplexes formed between DNAs carrying embryonic and plasmacytoma constant region gene sequences. *a*, A heteroduplex formed between λ Ch4A-EC-5 (embryonic) and λ Ch4A-41C-21 (plasmacytoma) DNA. *b*, A heteroduplex formed between Ch4A-EC-5 and Ch4A-149C-2 (plasmacytoma) DNA. The techniques for forming the heteroduplexes and spreading them for electron microscopy were as before¹⁴. The large arrow indicates the approximate location of the *Eco*R1 site in the phage DNA which is about 10.4 kilobases from the constant region gene (indicated by 'C'). The smaller arrows indicate the ends of single stranded regions which indicate non-homology between the two DNAs. The non-homology seen in these heteroduplexes is due to a DNA sequence present in plasmacytoma DNA, presumably introduced by a recombination event that occurred during the differentiation of the MOPC-149 and MOPC-41 cell lines. The measured lengths of single-stranded DNA in these heteroduplexes was used to determine the exact location of the recombination event with respect to ends of the *Eco*R1 fragments. Twenty-five molecules were measured.

DNA recrossed itself adjacent to the 5' end of the constant region R loop as shown by the molecule in Fig. 3a. That this structure is merely the result of the DNA twisting back over itself seems unlikely for three reasons. First, we never saw the DNA from the 10.4-kilobase side of the fragment twist over the end of the R loop. Second, in the 10 molecules observed, the J sequence was always 1.3 ± 0.19 kilobases from the end of the fragment. Third, we could not find this hairpin structure when we formed an R loop between MOPC-149 mRNA and the MOPC-149 DNA fragment and we knew from *in situ* hybridisation studies (Fig. 2) that this fragment does not have a J region sequence (Fig. 2). Thus, we conclude that the embryonic *Eco*R1 fragment contains a J region gene about 3.7 kilobases from the constant region gene. These features are summarised diagrammatically in Fig. 6.

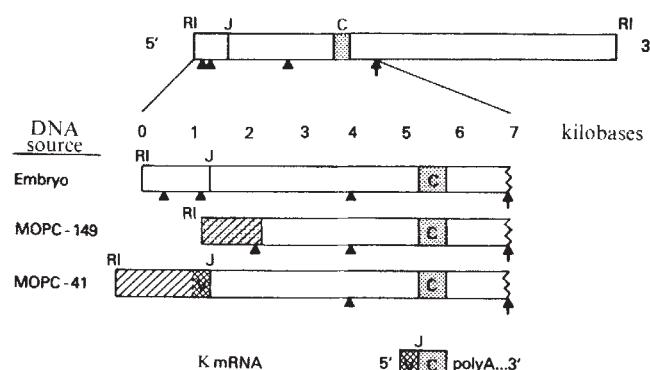


Fig. 6 Arrangement of light chain gene sequences on the *Eco*R1 fragments cloned from embryonic and plasmacytoma DNA. The *Eco*R1 fragment from embryo (16.0 kilobase) DNA is shown with the 5' end of the constant region gene to the left. The portions of the cloned fragments which differed from one another are shown on an expanded scale. The positions of the *Eco*R1, *Bam*H1 and *Hind*III restriction sites in this region are indicated. The segments of the light chain, the variable region (V), the J sequence (J) and the constant region (C) are indicated. The hatched segments in the plasmacytoma fragments, MOPC-149 and MOPC-41, represent DNA segments that are not present in the embryonic fragment. Note that they are different from each other as well. The κ light chain mRNA is diagrammed to indicate the size of the structural gene relative to the size of the intervening sequence in the MOPC-41 light chain gene.

The rearranged MOPC-41 fragment encodes a complete light chain gene and is interrupted by a 3.7-kilobase intervening sequence

In contrast to the embryonic and MOPC-149 segments, the 16.5-kilobase *Eco*R1 fragment cloned from MOPC-41 DNA hybridised to all three segments of the light chain sequence (Fig. 2). The R loop structure formed by annealing this fragment to MOPC-41 mRNA reflects this difference (Fig. 4). Instead of a single R loop, two are formed. Both flank a large, 3.7-kilobase intervening sequence that interrupts the structural gene and is not represented in the mature light chain mRNA. A tail, probably representing poly A, is frequently seen on the constant region end of the structure and its associated R loop measures 409 ± 128 base pairs. The variable region R loop measures 229 ± 68 base pairs. The sequence is also represented diagrammatically in Fig. 6.

Recombination sites defined on the two plasmacytoma-derived *Eco*R1 fragments

We have shown by restriction mapping and R loop analyses that the two plasmacytoma-derived fragments are related to, but different from each other and the fragment derived from embryonic DNA. In order to define more closely the relationship between these three constant region encoding fragments and to compare their sequences, we have formed heteroduplexes between these fragments cloned in the λ Ch4A vector. An example of the heteroduplex structures formed between the phage containing the embryonic fragment and the phage containing the MOPC-41 segment is shown in Fig. 5a. The genomic fragments (whose ends are defined by the arrows closest to the ends of the double-stranded DNA in Fig. 5a) form a homologous, double-stranded structure for about 14.8 kilobases and then diverge at one end (the single-stranded bubble). Since the right arm of the λ Ch4A vector DNA is shorter than the left¹⁶, the orientation of the fragment and the location of the differences between the two fragments are defined at the points of divergence in the heteroduplex molecules. Thus, the rearranged MOPC-41 fragment differs from embryonic fragments by a segment of DNA introduced at its 5' end at approximately the position of the embryonic J sequence (Fig. 6).

Similarly, the point of divergence in the rearranged MOPC-149 fragment could be determined. A heteroduplex formed between the phages containing the embryonic fragment and the

MOPC-149 fragment is shown in Fig. 5b. The fragments are homologous for 13.5 kilobases, but also diverge at their 5' end. In each case, the divergent heteroduplex bubble represents DNA that is different in the embryonic sequence. Evidently these new sequences are derived from a recombination-like event which joined DNA sequences to the 5' ends of plasmacytoma DNA fragments and eliminated sequences present in the embryonic fragment (see diagram, Fig. 6). By measuring the lengths of the single-stranded DNAs found in each heteroduplex, we can estimate the lengths of the new DNA sequences that have been joined to the embryonic *Eco*R1 fragment. The points at which the embryonic fragment differs from each rearranged fragment defines the site at which the recombination event took place.

The recombination events that formed the new MOPC-149 and MOPC-41 fragments involved different sites. In MOPC-149 the recombination event appears to have occurred 2.7 kilobases from the constant region gene and removed the MOPC-41-like J region sequence from the embryonic fragment. Presumably the MOPC-149 light chain uses a different J region sequence. The recombination event that generated the MOPC-41 light chain sequence occurred 3.7 kilobases from the constant region gene, very close to the J region sequence found in the embryonic fragment. Here a single recombination event may have linked the MOPC-41 variable region to the MOPC-41 J region to generate the cloned fragment. However, until the nucleotide sequences of the appropriate regions have been determined, we cannot be certain that the recombination event that generated the MOPC-41 gene did not actually replace the embryonic J region with a MOPC-41 J region at the same relative position. It is also possible that both J region sequences are present on the MOPC-41-derived fragment. Nevertheless, these analyses indicate that the recombination events that affect immunoglobulin light chain genes can involve sites at least 1.0 kilobases apart.

The emerging picture of antibody gene arrangement and rearrangement

By characterising cloned, κ chain-encoding segments of DNA derived from mouse embryo and antibody producing cells, we demonstrate directly that immunoglobulin constant and variable region genes undergo a somatic rearrangement that brings together the segments necessary to encode a complete κ light chain gene. Moreover, in each light chain producing plasmacytoma that we have thus far examined there is evidence of somatic rearrangement of one set of these genes, while a second, presumably allelic, set retains the original embryonic pattern. These observations fit well with the plausible notion that rearrangement of a single allele is instrumental in immunoglobulin gene activation and that it may play a part in allelic exclusion in antibody production.

In addition we see that, like the mouse λ light chain⁹, κ genes are encoded in at least three segments in embryonic DNA, a foreshortened variable region sequence⁴, a short J sequence probably encoding the last 10 or so amino acids of the variable region and, finally, the constant region sequence. Rearrangement of these sequences (which may itself contribute to the generation of antibody diversity) forms, in the case of the MOPC-41 light chain gene, a new chromosomal sequence in which variable and J sequences are adjacent, but separated from the constant region gene by a large 3.7-kilobase intervening sequence of DNA. In the case of MOPC-149, this rearrangement has eliminated the embryonic J sequence and replaced it, in our clone, with a new segment of DNA. Whether this new segment is a portion of the intervening sequence linking MOPC-149 variable and constant regions in genomic DNA is not known. In contrast to what might have been expected, these rearrangements do not occur at a single site, but the new DNA segments are joined at sites about 1.0 kilobases apart on the 5' side of the constant region gene. The precise nature and the molecular basis of this recombination remains unclear. Further studies should indicate whether or not reciprocal recombination is involved and whether this recombination is mediated by regions of sequence homology.

We thank Dr Fred Blattner for the λ CH4A phage system, Barbara Norman, Marion Nau and Aya Leder for help in the gene purification and cloning experiments, Margery Sullivan for help with the electron microscopy and Dr Henry I. Miller for help with preparative gel electrophoresis. We thank Aya Leder for suggesting the gene orientation experiments shown in Fig. 3b, and Terri Broderick for assistance in preparation of this manuscript.

Received 17 October; accepted 27 November 1978.

1. Dreyer, W. J. & Bennett, J. C. *Proc. natn. Acad. Sci. U.S.A.* **54**, 864-868 (1965).
2. Tonegawa, S., Hozumi, N., Matthysens, G. & Schuller, R. *Cold Spring Harb. Symp. quant. Biol.* **41**, 877-889 (1976).
3. Tonegawa, S., Brack, C., Hozumi, N. & Schuller, R. *Proc. natn. Acad. Sci. U.S.A.* **8**, 3518-3522 (1977).
4. Seidman, J. G. *et al. Proc. natn. Acad. Sci. U.S.A.* **75**, 3881-3885 (1978).
5. Hozumi, N. & Tonegawa, S. *Proc. natn. Acad. Sci. U.S.A.* **73**, 3628-3632 (1976).
6. Tonegawa, S., Brack, C., Hozumi, N. & Pirrotta, V. *Cold Spring Harb. Symp. quant. Biol.* **42**, 921-931 (1977).
7. Rabbitts, T. H. & Forster, A. *Cell* **13**, 319-327 (1978).
8. Tonegawa, S., Maxam, A. M., Tizard, R., Bernard, O. & Gilbert, W. *Proc. natn. Acad. Sci. U.S.A.* **75**, 1485-1489 (1978).
9. Brack, C., Hirama, M., Lenhard-Schuller, R. & Tonegawa, S. *Cell* **15**, 1-14 (1978).
10. Kabat, E. A., Wu, T. T. & Bilofsky, H. *Proc. natn. Acad. Sci. U.S.A.* **75**, 2429-2433 (1978).
11. Seidman, J. G., Leder, A., Nau, M., Norman, B. & Leder, P. *Science* **202**, 11-17 (1978).
12. Southern, E. M. *J. molec. Biol.* **98**, 503-517 (1975).
13. Tilghman, S. M. *et al. Proc. natn. Acad. Sci. U.S.A.* **10**, 4406-4410 (1977).
14. Tiemeier, D. C. *et al. Cell* **14**, 1309-1313 (1978).
15. Leder, A. *et al. Proc. natn. Acad. Sci. U.S.A.* (in the press).
16. Blattner, F. R. *et al. Science* **196**, 161-169 (1977).
17. Bek, A. & Gould, M. *Proc. natn. Acad. Sci. U.S.A.* **72**, 581-585 (1975).
18. Sternberg, N., Tiemeier, D. & Enquist, L. W. *Gene* **1**, 255-280 (1977).
19. Seidman, J. G., Edgell, M. H. & Leder, P. *Nature* **271**, 582-585 (1978).
20. White, R. L. & Hogness, D. S. *Cell* **10**, 177-192 (1977).
21. Tilghman, S. M. *et al. Proc. natn. Acad. Sci. U.S.A.* **75**, 725-729 (1978).
22. Lenhard-Schuller, R. *et al. Proc. natn. Acad. Sci. U.S.A.* **75**, 4709-4713 (1978).

Synthesis of growth hormone by bacteria

P. H. Seeburg, J. Shine, J. A. Martial, R. D. Ivarie, J. A. Morris, A. Ullrich, J. D. Baxter & H. M. Goodman

Howard Hughes Medical Institute Laboratory, Department of Biochemistry and Biophysics, Department of Medicine, and Metabolic Research Unit, University of California, San Francisco, California 94143

A hybrid gene was constructed between the β -lactamase gene of plasmid pBR322 and the cloned coding sequence for rat growth hormone. This gene is expressed in bacteria and growth hormone sequences are detectable by immunological methods.

RECENT advances in recombinant DNA technology have led to the cloning in bacteria of the natural coding sequences for several mammalian peptide hormones¹⁻³. These sequences provide excellent material for testing the feasibility of producing mammalian proteins in bacteria. The success of such production depends on the correct design and construction of new genes in which suitable bacterial control elements promote expression of

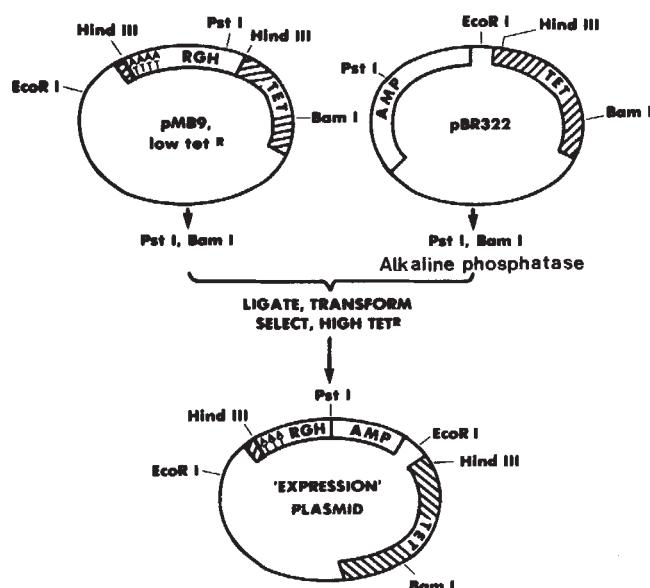


Fig. 1 Construction of a vector for bacterial synthesis of GH sequences. The cloned RGH cDNA was cleaved from plasmid pRGH-1 (ref. 2) with *Hind*III endonuclease and isolated by polyacrylamide gel electrophoresis. This DNA was enzymatically ligated to similarly cleaved and alkaline phosphatase-treated^{1,3} pMB9. The ligation mixture was used to transform *E. coli* strain χ 1776 (refs 1–3). Colonies containing plasmids were selected by growth on low levels ($5 \mu\text{g ml}^{-1}$) of tetracycline¹¹. Resistant colonies were examined for the presence of plasmids containing inserted RGH sequences by gel analysis of such plasmids after digestion with restriction endonuclease *Pst*I and *Eco*RI. In some plasmids RGH DNA was found to have its poly dA·dT end orientated towards the single *Eco*RI site in pMB9. Such a plasmid (pMB9-RGH) was used in the subsequent steps for vector construction. The pre-RGH sequences from pMB9 were placed under the control of the β -lactamase gene of pBR322, as indicated. Plasmid pMB9-RGH was digested with *Pst*I plus *Bam*HI endonuclease. The single cleavage site for *Pst*I in this plasmid occurs between amino acids –24 and –23 of the pre-sequence of RGH (see Fig. 2), whereas the single *Bam*HI site is in the gene responsible for bacterial tetracycline resistance⁹. The *amp*^r and *tet*^r plasmid pBR322 was similarly digested with these two enzymes. In pBR322 the *Pst*I site occurs at a sequence coding for amino acids 182 and 183 of pre- β -lactamase¹⁰. The *Bam*HI site is in the *tet*^r gene and is in exactly the same place as in pMB9, as this region of pBR322 was derived from pMB9 (ref. 9). After digestion, pBR322 was treated with bacterial alkaline phosphatase^{1,3}. This mixture was then ligated to the *Pst*I and *Bam*HI cleaved pMB9-RGH. After transformation of *E. coli*, colonies were selected for their ability to grow on high levels ($20 \mu\text{g ml}^{-1}$) of tetracycline.

the appropriate coding sequences. We describe here the construction of such a hybrid gene on a bacterial plasmid designed to program bacteria for the synthesis of pituitary growth hormone (GH) sequences. This peptide hormone is crucial for linear growth. A sufficient supply would allow wider therapeutic use in the treatment of growth disorders^{4–6} and other conditions in man and might be useful in animal husbandry for more economical food production. The species specificity and size (190 amino acids) of GH^{7,8} prevents us from exploiting existing resources or applying current synthetic techniques. Thus, for GH and similarly for a variety of other eukaryotic proteins, genetically programmed bacteria may provide an important resource.

Gene construction

The construction of a hybrid gene between bacterial and mammalian DNA sequences, capable of directing synthesis of rat

Table 1 Radioimmunoassay of RGH in extracts from minicells containing pEx-RGH or pBR322

Plasmid	RGH by radioimmunoassay (ng ml^{-1})	
	Spheroplasts	Periplasmic space
pEx-RGH	89.1 (70.8–116.6)	30.8 (30.0, 31.6)
pBR322	Undetectable	Undetectable

One litre of *E. coli* P648-54 containing either pEx-RGH or pBR322 was grown to saturation and collected by centrifugation at $10,000g$ for 10 min. The cells were washed and spheroplasts were prepared by modification of procedures described elsewhere¹⁵. Briefly, cells were washed once in 200 ml of 10 mM Tris (pH 7.4) containing 1 mM PMSF (phenylmethylsulphonyl fluoride, Sigma). The cells were resuspended in 50 ml of 20% sucrose, 0.033 Tris (pH 8.0) and 1 mM PMSF. Ice-cold 0.1 M EDTA (pH 8.0) was added (0.5 ml) and the mixtures were kept on ice for 5 min. After addition of 1 ml of 5 mg ml^{-1} lysozyme (Sigma) the mixtures were incubated for 30 min at 0°C and centrifuged ($17,000g$, 10 min, 0°C). The supernatant media were saved for radioimmunoassay of material from the periplasmic space. The pellets were resuspended in 10 mM Tris (pH 8.0) and 1 mM PMSF. MgCl_2 was added to a final concentration of 10 mM with 200 μl each of DNase (1 mg ml^{-1}) and RNase (5 mg ml^{-1}). Incubation proceeded for 1 h at 4°C (spheroplast sample). These mixtures (10 ml) and 50 ml of the samples containing periplasmic material were precipitated with $(\text{NH}_4)_2\text{SO}_4$ (50% saturation). The samples were resuspended in 4 ml of 50 mM Tris (pH 9.0), 0.1 M NaCl and 1 mM PMSF and dialysed against 2 l of the same buffer for 2 h at 4°C . Serial dilutions of these samples were analysed by radioimmunoassay as described previously¹⁹. Amounts of RGH found by this assay are listed in ng ml^{-1} with the mean and range of these determinations for the spheroplast sample and the mean and individual values for duplicate samples from the periplasmic space sample. No activity was detected in samples from pBR322. The sensitivity minimum of the assay as assessed from the standard curve with authentic GH was 10 ng ml^{-1} . The high immune reactivity in the periplasmic space is probably due to lysis of some cells during spheroplast formation.

growth hormone (RGH) sequences in bacteria, is shown in Fig. 1. The bacterial sequences for this gene were excised from the multicopy plasmid pBR322 (ref. 9). They contain a portion of the gene (*amp*^r) coding for the enzyme β -lactamase which is secreted into the periplasmic space and is responsible for bacterial resistance to the antibiotic ampicillin⁹. The mammalian sequences were provided by cloned complementary DNA containing the entire coding region for the precursor form of rat pituitary GH². Functional fusion of the two donor DNAs was facilitated by knowledge of their primary structures^{2,10}. Both DNAs possess a unique cleavage site for the restriction endonuclease *Pst*I at which they can be joined to maintain the correct reading frame (Fig. 2). In the *amp*^r gene the *Pst*I site is situated approximately two-thirds towards the end. In RGH cDNA the corresponding site occurs a few base pairs distal to the

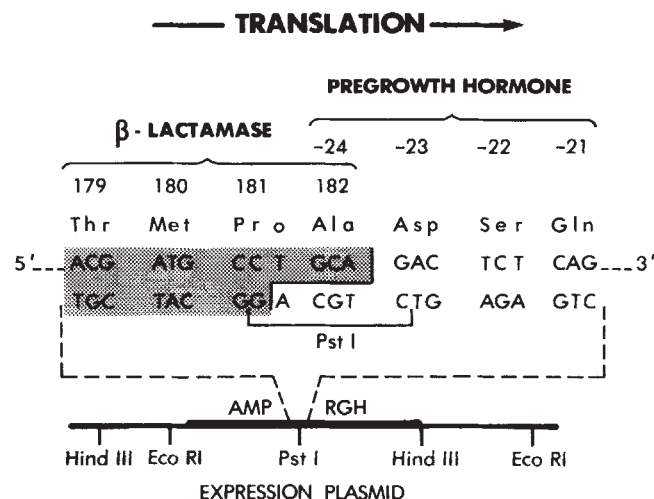


Fig. 2 Postulated nucleotide sequence around the *Pst*I site in the hybrid gene. The *Pst*I site is at the fusion between the *amp*^r gene sequences¹⁰ and the cloned RGH cDNA². Positive numbers refer to amino acid residues of pre- β -lactamase and negative numbers to those of the RGH pre-sequence.

initiator codon for the pre-peptide. The resultant new gene should code for a chimaeric protein of 395 amino acids (molecular weight $\sim 44,000$) containing the N-terminal 181 amino acids of the β -lactamase precursor covalently linked to 214 residues of rat pre-GH. As this fusion protein carries the signal peptide of β -lactamase, it seems possible that it might be transported to the periplasmic space. Such an extracellular location might minimise possible protease attack and facilitate isolation.

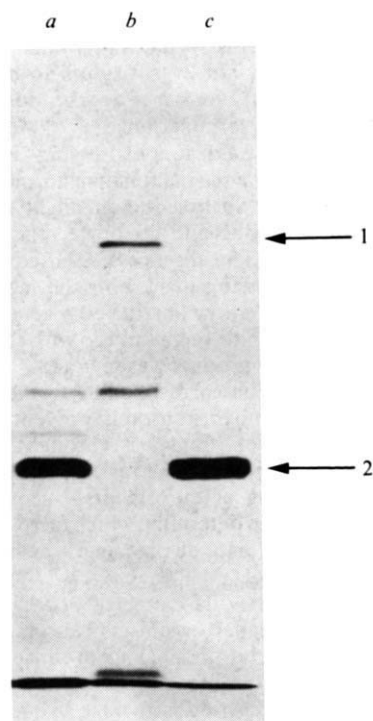


Fig. 3 Proteins synthesised from plasmids pEx-RGH and pBR322. The minicell-producing *E. coli* strain P678-54 (ref. 13) was transformed with pEx-RGH or pBR322 by standard methods². A culture of each (200 ml) was grown to the late log or stationary phase of growth ($A_{650} = 0.7$) in L broth containing 0.5% glucose. Minicells were purified by repeated zone sedimentation through 5–25% sucrose gradients as described elsewhere¹² and resuspended in 2 ml of leucine- and methionine-free M-9 salts medium. After 30 min preincubation at 37 °C, 250 μ Ci of 35 S-methionine and 50 μ Ci of 3 H-leucine were added to both cell preparations, which were then vigorously shaken for 30 min at 37 °C. After a 30-min chase period with 400 μ g of nonradioactive methionine and leucine, the minicells were pelleted by centrifugation at 8,000g for 2 min and half the cells of each preparation were lysed in 200 μ l of SDS sample buffer¹⁴. The second half of each minicell preparation was used to isolate the proteins secreted into the periplasmic space as described elsewhere¹⁵. Labelled proteins isolated from equivalent amounts of minicells were electrophoresed on a 10% polyacrylamide-SDS gel¹⁴ and identified by autoradiography of the dried gels. Lanes a and b represent the total proteins synthesised by pBR322 and pEx-RGH, respectively. Lane c reflects the proteins secreted into the periplasmic space of minicells carrying pBR322. Arrow 1 refers to the location of the fusion protein and arrow 2 indicates β -lactamase.

To enable selection of plasmids carrying the correct hybrid gene, the cloned GH cDNA (originally isolated in plasmid pBR322) was first transferred from the *Hind*III site of pBR322 into the homologous site of pMB9, a closely related plasmid which lacks the *amp*^r gene. In the resultant recombinant plasmid (pMB9-RGH) the gene for tetracycline resistance is expressed at reduced levels¹¹. One such plasmid which contains the GH cDNA inserted in the orientation shown in Fig. 1 was used as the recipient of *amp*^r gene sequences. Replacement of the unique

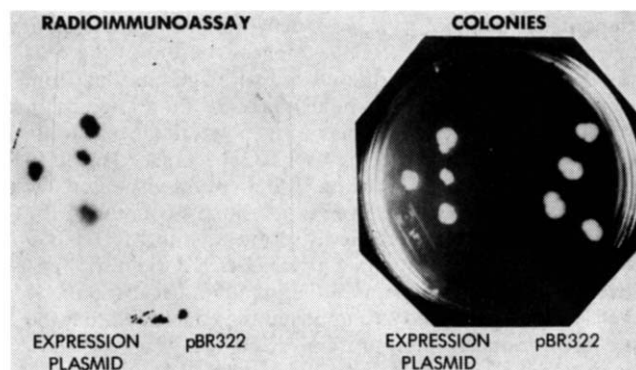


Fig. 4 Immunological detection of RGH sequences in bacteria. Bacterial colonies containing the constructed expression vector pEx-RGH were assayed for the presence of RGH sequences using a modification¹⁶ of the solid-phase immunological screening method Broome and Gilbert¹⁷. The described wash buffer was replaced with phosphate-buffered saline (0.1 M NaCl, 0.025 M potassium phosphate, pH 7.4) containing gelatin (10 mg ml⁻¹) and NP40 (0.1%). Bacterial colonies of *E. coli* strain P 678-54 (ref. 13) carrying pEx-RGH or pBR322 were grown on L-broth agar¹⁸ in Petri dishes. A photograph of such a plate is shown on the right. After treatment (20 min) with chloroform vapour the colonies were exposed to polyvinyl strips coated with antiserum to RGH. Strips were then soaked in the modified wash buffer (see above) to minimise binding of IgG in the subsequent reaction. The strips were then incubated overnight at 4 °C with affinity column-purified 125 I-anti-RGH IgG (10⁶ c.p.m., specific activity ~ 5.3 μ Ci per μ g), washed, and exposed to X-ray film. An autoradiogram is shown on the left.

*Pst*I–*Bam*HI fragment in pMB9-RGH with that from pBR322 should place the coding sequences for rat GH under control of the *amp*^r gene and will at the same time restore a fully functional *tet*^r gene (see Fig. 1). Thus, this fragment replacement enabled selection of bacterial colonies carrying the expression vector 'pEx-RGH' by their ability to grow on high levels of tetracycline.

Gene expression

Purified minicells¹² were used for the study of proteins encoded by the expression vector pEx-RGH. Proteins were analysed by SDS-polyacrylamide gel electrophoresis and compared with those encoded by pBR322 (Fig. 3) and pMB9 (data not shown). As anticipated, the two polypeptides pre- β -lactamase and β -lactamase which are synthesised in cells carrying pBR322 are not made from pEx-RGH. The expression vector codes uniquely for a polypeptide (Fig. 3, lane b, arrow 1) with an apparent molecular weight (46,000) within error limits of that expected for the hybrid protein. A comparison of band intensities (corrected for cell number) from several experiments similar to that described in Fig. 3 suggests that the 46,000 MW protein is produced in substantial quantities from pEx-RGH, although in lower amounts (about one-fifth) than β -lactamase from pBR322. When proteins released after the formation of spheroplasts in high osmotic pressure medium¹⁵ were analysed by gel electrophoresis, a large portion of β -lactamase (Fig. 3) but not of the pEx-RGH specific product was detected. The same situation applied when material from media of whole cells was examined (data not shown). Thus, contrary to the case with β -lactamase, the 46,000 MW protein was present only in very small amounts in the periplasmic space if at all.

Immunological detection of growth hormone synthesised by bacteria

Escherichia coli colonies containing pEx-RGH were screened for their content of RGH using a modification¹⁶ of a solid-phase

immunological method¹⁷ (Fig. 4). Briefly, colonies were lysed *in situ* with CHCl_3 and any RGH sequences present in the cells were bound to a polyvinyl disk coated with IgG purified from antiserum to RGH raised in monkeys (see ref. 16 for details on preparation of antiserum). Bound antigen was then specifically labelled with highly purified ^{125}I -anti-RGH IgG (see legend to Fig. 5) and colonies producing RGH were visualised by autoradiography. As shown in Fig. 4, colonies containing the expression plasmid were labelled with the ^{125}I -anti-RGH-IgG probe, whereas those containing pBR322 failed to do so.

Several controls demonstrated the immunological specificity of the assay for RGH and strengthened the conclusion that the clones are producing pituitary GH sequences (Fig. 5). First, colonies of the K12 strain of *E. coli* devoid of plasmid were not labelled with the ^{125}I -anti-RGH-IgG probe. Second, labelling of lysates of colonies containing pEx-RGH was prevented when a large excess (250-fold) of nonradioactive GH was incubated with radioactive probe. Third, labelling was not observed when normal monkey serum or calf serum (data not shown) were used in place of anti-RGH antiserum. Finally, ^{125}I -labelled normal monkey serum failed to label any colony from the strains described above (not shown).

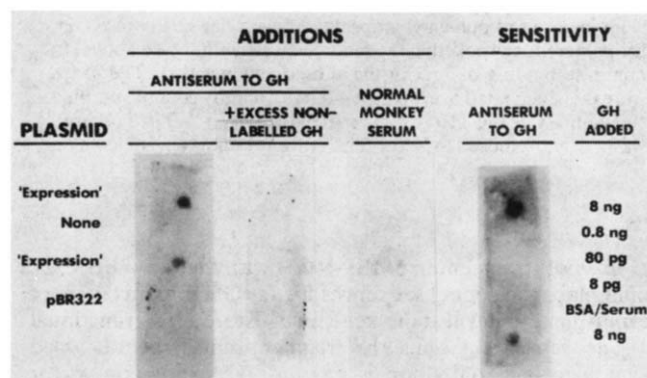


Fig. 5 Immunological detection of GH sequences synthesised by bacteria: specificity of the assay. In 'antiserum to GH', the colonies were treated as described in the legend to Fig. 4. 'None' refers to colonies of *E. coli* devoid of plasmid. In other experiments ('+ excess non-labelled GH'), a large excess (250-fold, 5 μg) of highly purified RGH was preincubated with ^{125}I -anti-RGH IgG. The autoradiograph shown in the lane 'normal monkey serum' results from using polyvinyl strips coated with normal monkey serum (in place of RGH antiserum). A slight reaction occurred with cells containing pEx-RGH, as antigen (RGH) sticks to plastic even in the absence of anti-RGH IgG. The sensitivity of the assay was measured by applying varying amounts of pure RGH (8 ng, 0.8 ng, 80 pg, 8 pg and 8 ng) to an agar plate and screening as described above.

The sensitivity of the screening procedure is indicated in Fig. 5. Labelling was observed with 8 ng, but not with 0.8 ng, of authentic RGH added to the plate. Thus, each colony seems to contain several nanograms of GH sequences, although further experiments are needed to obtain better quantitation.

As indicated in Table 1, immune reactivity was also detected by direct radioimmunoassay in extracts of spheroplasts and in material from the periplasmic space. No such activity was observed in similar isolates from bacteria containing pBR322.

Discussion

In the present studies, bacteria have been genetically programmed to synthesise mammalian pituitary growth hormone sequences. Further analysis is required to assess the fidelity of

such expression, the stability of the protein products, and their biological activity. The absolute quantity of GH synthesised by the bacteria cannot be determined accurately from the current studies (an estimate of 24,000 molecules per cell is obtained from the solid-phase immunoassay). However, judging from the prominence of the band in Fig. 3, it seems that sequences of RGH can be a major plasmid product. The quantity of GH present is less than that of β -lactamase; however, the study shown in Fig. 3 may result in an underestimate of the amount of GH synthesised, because in more recent studies (not shown), material of lower molecular weight (in addition to the larger molecular weight material) is specifically immunoprecipitated by antiserum to RGH but not by antiserum to bovine serum albumin. This material could consist of degradation products or the products of premature termination.

The *amp^r* gene was used for expressing RGH coding sequences with the hope that the mammalian hormone would be secreted by the bacteria. Such secretion might facilitate the isolation of the protein product and also protect the protein from degradation by proteases. The absence of GH sequences in the periplasmic space is surprising in view of the fact that Villa-Komaroff *et al.*²⁰ have recently constructed a gene suitable for the expression of rat insulin through linkage with the *amp^r* gene and found that the protein product was secreted²⁰. It is possible that the hydrophobic region corresponding to the pre-part of RGH in the middle of the hybrid protein prevents the kind of conformation required for transport through membranes. Similarly, the overall structure of the fusion polypeptide might prevent correct processing of the precursor sequence. More detailed studies should help determine which factors, other than pre-sequence of the signal peptide protein of the hormone, are necessary for secretion.

Our findings and those of Villa-Komaroff *et al.*²⁰ indicate that coding sequences from higher organisms can be expressed in bacteria. Thus, it should be possible to produce biologically important peptides with the use of recombinant DNA techniques and naturally occurring structural gene sequences. Finally, the capability of expression of genes that are designed specifically or synthesised²¹ should make possible the production of new hormone analogues for examining the structural requirements for agonist or antagonist activity or for therapy.

H.M.G. and J.D.B. are Investigators of the Howard Hughes Medical Institute. We thank Edmund Tisher for help with these studies. This work was supported in part by NIH grants AM-19997, CA14026 and CA23046, NSF grant PCM 77-16453, and a grant from Lilly.

Received 20 October; accepted 13 November 1978.

- Ullrich, A. *et al. Science* **196**, 1313-1319 (1977).
- Seeburg, P. H., Shine, J., Martial, J. A., Baxter, J. D. & Goodman, H. M. *Nature* **270**, 486-494 (1977).
- Shine, J., Seeburg, P. H., Martial, J. A., Baxter, J. D. & Goodman, H. M. *Nature* **270**, 494-499 (1977).
- Goodman, H. G., Grumbach, M. M. & Kaplan, S. L. *New Eng. J. Med.* **278**, 57-68 (1968).
- Tanner, J. M., Whitehouse, R. H., Hughes, P. R. C. & Vince, F. P. *Archs Dis. Child.* **46**, 745-782 (1971).
- Winawa, S. J., Sherlock, P., Lipkin, M., Sonenberg, M. & Vanamer, P. *Advances in Human Growth Hormone Research*, 862-874, DHEW Publ. No. (NIH) 74-612 (1973).
- Wallis, M. & Davies, R. V. in *Growth Hormone and Related Polypeptides* (eds Pecile, A. & Muller, E. E.) 1-14 (Elsevier, New York, 1976).
- Knobil, E. & Greer, R. D. *Recent Prog. Horm. Res.* **15**, 1-12 (1959).
- Bolivar, F., Rodriguez, Betlach, M. C. & Boyer, H. W. *Gene* **2**, 75-93 (1977).
- Sutcliffe, J. G. *Proc. natn. Acad. Sci. U.S.A.* **75**, 3737-3741 (1978).
- Rodriguez, R. L., Bolivar, F., Goodman, H. M., Boyer, H. W. & Betlach, M. in *ICN/UCLA Symp. Molec. Mech. Control Gene Expression* (eds Nierlich, D. P., Rutter, W. J. & Fox, C. F.) 471-477 (Academic, New York, 1976).
- Roozen, K. J., Fenwick, R. G. Jr & Curtis, R., III *J. Bact.* **107**, 21-33 (1971).
- Adler, H. I., Fischer, W. D., Cohen, A. & Hardigier, A. A. *Proc. natn. Acad. Sci. U.S.A.* **57**, 321-326 (1966).
- Leammli, U. K. *Nature* **227**, 680-685 (1970).
- Neu, H. C. & Hepel, L. A. *J. biol. Chem.* **240**, 3685-3692 (1965).
- Ivarie, R. D. & Morris, J. A. (in preparation).
- Broome, S. & Gilbert, W. *Proc. natn. Acad. Sci. U.S.A.* **75**, 2746-2749 (1978).
- Miller, J. H. *Experiments in Molecular Genetics* (Cold Spring Harbor Laboratory, New York, 1976).
- Papavasiliou, S. S., Martial, J. A., Latham, K. R. & Baxter, J. D. *J. clin. Invest.* **60**, 1230-1239 (1977).
- Villa-Komaroff, L. *et al. Proc. natn. Acad. Sci. U.S.A.* **75**, 3727-3731 (1978).
- Itekura, K. *et al. Science* **198**, 1056-1063 (1977).

letters

Discovery and optical identification of 2S0921-630

THE discovery of a faint X-ray source with the X-ray detectors on SAS 3 at a location with galactic coordinates $l^{\text{II}} \sim 282^\circ$, $b^{\text{II}} \sim -9^\circ$ is reported here. A subsequent optical search has revealed a likely optical counterpart in the form of a 17th magnitude star with He II 4686 and H β in emission. It is, therefore, likely to be a low-luminosity galactic X-ray source.

From 10 to 14 February 1978, the rotation modulation collimator X-ray detectors¹ on SAS 3 were used to search for X-ray emission from a region centred at $\alpha \sim 135^\circ$, $\delta \sim -67^\circ$. Data were accumulated from a total exposure of 130,000 s. The correlation maps show a previously unreported X-ray source, which we designate 2S0921-630 at $\alpha = 09 \text{ h } 21 \text{ min } 25.4 \text{ s}$, $\delta = -63^\circ 04' 27''$ (1950 Equinox) with an error circle of radius $30''$ (90% confidence). Figure 1 is an enlarged reproduction from the ESO (B) survey print with the error circle superposed. The average flux density of the source during the observation was $2.3 \pm 0.3 \mu\text{Jy}$ at 5.2 keV ($1 \mu\text{Jy} = 0.242 \times 10^{-11} \text{ erg cm}^{-2} \text{ s}^{-1} \text{ keV}^{-1}$).

Optical observations of stars in the above error circle were made with the 3.6 m telescope of the European Southern Observatory (ESO). To search for a candidate counterpart with possible strong H α emission, one of us (S.L.) took a direct prime-focus plate, through a $30 \text{ \AA}$ wide H α filter, of the region of the sky centred on the error box. Star No. 5 in Fig. 1 is remarkably bright on this plate. We estimate from the ESO survey print that its photographic magnitude is approximately 17.

Spectrograms ($60 \text{ \AA mm}^{-1}$) were taken (by J.v.P.) of this star with the Carnegie image tube attached to the Boller and Chivens Cassegrain spectrograph. The plates were taken on 13 March 1978, 03.33 UT and 04.07 UT and have exposure times of 40 and 20 min, respectively. H α -O (baked) plates were used and

the spectra have a widening of 0.2 mm. They cover the wavelength range from $\sim 3,700$ to $\sim 5,000 \text{ \AA}$. Density tracings of the spectra were made with the Faul-Coradi microphotometer of the Observatory of Utrecht. The only features clearly visible in the spectrum are a strong He II 4686 and a weak H β line, both in emission. Strong He II emission lines are generally present in the spectra of the counterparts of X-ray sources (see refs 2 and 3) but not usually in those of normal stars; on the basis of the above characteristics we consider star No. 5 a likely candidate for the optical counterpart of 2S0921-630.

The optical spectrum of this object is similar to the spectra of the optical counterparts of Sco X-1 and the similar galactic X-ray sources that have been discussed by McClintock *et al.*². It is interesting, however, that the ratio of X-ray to optical luminosity for 2S0921-630 is ~ 10 which is much smaller than the ratios for other members of this class of counterparts which are typically of the order of 1,000. The X-ray luminosity of the sources is $\sim 6 \times 10^{35} (D/10 \text{ kpc})^2 \text{ erg s}^{-1}$. It may belong to the class of low luminosity galactic X-ray binaries.

We thank the staff at MIT Center for Space Research for help with this observation. Permission to use the Faul-Coradi microphotometer is gratefully acknowledged.

This work was supported in part by the NASA. J.A.v.P. is on leave from the Astronomical Institute, University of Amsterdam.

F. K. LI
J. A. VAN PARADIJS
G. W. CLARK
J. G. JERNIGAN

Department of Physics and Center for Space Research,
Massachusetts Institute of Technology,
Cambridge, Massachusetts 02139

S. LAUSTSEN

European Southern Observatory,
Casilla 16317, Santiago, Chile

E. J. ZUIDERWIJK

Astronomical Institute, University of Amsterdam,
Roetersstraat 15, Amsterdam, The Netherlands

Received 7 August; accepted 27 October 1978.

1. Doxsey, R. *et al. Astrophys. J. Lett.* **203**, L9-L13 (1976).
2. McClintock, J. E., Canizares, C. R. & Backman, D. *Astrophys. J. Lett.* **223**, L75 (1978).
3. McClintock, J. E., Canizares, C. R. & Tarter, C. B. *Astrophys. J.* **198**, 641-652 (1975).

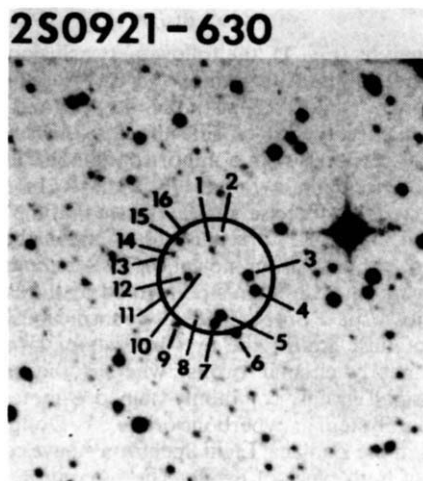


Fig. 1 Error box of 2S0921-630 superposed on the ESO survey print. Star No. 5 is the proposed optical counterpart.

Simultaneous radio and X-ray observations of MXB1837+05 (Ser X-1)

DURING a continuing program of radio observations of X-ray sources carried out with the NRAO very large array (VLA), we have made high sensitivity searches for radio emission from MXB1837+05 = Ser X-1 (ref. 1), particularly during the SAS 3 MIT world-wide X-ray burster watch in June 1977². Observations of MXB1837+05 were carried out on 26 April, 12 and

13 June 1977, with six 25-m antennas of the VLA operated at 4,884 MHz with system temperatures of ~ 60 K and bandwidths of 50 MHz. Upper limits of 1 mJy were obtained for the time-averaged radio emission from Ser X-1 during the three periods of observation, and no short time scale radio emission was detected during or after the times X-ray bursts were observed on 12 and 13 June 1977.

The times of observations with the VLA were from 06.01 UT to 13.50 UT on 26 April 1977, from 06.15 to 13.18 UT on 12 June and from 07.04 UT to 13.14 UT on 13 June 1977. During these periods calibration sources were observed once every half hour. The VLA data were accumulated in 30-s integration intervals, and after calibration, were used to make radio maps of about 4 arc min square regions with 2 arc s resolution. The centre of each map was taken to be the position of MXB1837+05 ($\alpha(1950) = 18^{\text{h}}37^{\text{m}}29.8^{\text{s}}$, $\delta(1950) = +04^{\circ}59'23''$) which was known to about 20" (ref. 3). Maps were made from various subsets of each 7-h period allowing upper limits to be set for variable radio emission on time scales from 30 s up to ~ 7 h. On each of the three days the upper limits to the time-averaged radio emission anywhere in the fields of the map were ~ 1 mJy. The times of major interest were the periods during and after X-ray bursts observed by SAS 3 on 12 and 13 June. The upper limits placed on radio emission during and after the bursts are a factor of 8 lower than the limits established by Johnson *et al.*⁴ for X-ray bursters including MXB1837+05 and a factor of 100 lower than the limits found by Thomas *et al.* for MXB1636-53.

The X-ray observations were made in the 1.3-19-keV energy range with the SAS 3 horizontal tubes and xenon tube system^{6,7}. Two X-ray bursts were observed during VLA radio coverage on 12 June at 11.22.20 and 13 June at 9.29.27 UT. The X-ray bursts lasted ~ 17 s with average X-ray fluxes of 0.2-0.4 mJy. Maps with integration times of 30s, 10 min, 2 h, and 7 h were made from the radio data, with upper limits for any sources in the field of 15, 4, 1.5 and 1.0 mJy, respectively, with the shorter time periods taken coincident with or after the X-ray bursts. Exact simultaneity was achieved with the 12 June burst at 11.22.20 UT, but the 13 June burst at 9.29.27 UT occurred during a 5-min gap when a calibrator was being observed. With the limits mentioned above for the four time scales of mapping, no radio emission was observed during or after the 12 June burst and, no radio emission was observed before or after the 13 June burst.

In conclusion, even very high sensitivity radio observations cannot detect bursting X-ray sources. A clearer picture of these objects is needed, however, before the radio observations can effectively be discussed on a theoretical basis.

The MIT research effort was supported in part by NASA contract NAS5-11450. M.P.U., supported in part by NASA grant NAG5196, thanks MIT for support and hospitality during part of this work. We thank Christian Tellefson, Steve Holzner, Bob Dobbin, and the SAS 3 group for invaluable help.

M. P. ULMER

Department of Physics and Astronomy,
Northwestern University,
Evanston, Illinois 60201

R. M. HJELLMING

National Radio Astronomy Observatory,
Socorro, New Mexico 87801

W. H. G. LEWIN
J. A. HOFFMAN
J. G. JERNIGAN
W. WHEATON
F. PRIMIMI
J. DOTY
H. MARSHALL

Department of Physics and
Center for Space Research,
Massachusetts Institute of Technology,
Cambridge, Massachusetts 02139

Received 2 August; accepted 6 September 1978.

1. Li, F. L. *et al. Mon. Not. R. astr. Soc.* **179**, 21P-25P (1977).
2. Lewin, W. H. G. *IAU Circ.* No. 3078 (1977).
3. Doxsey, R. E., Apparo, K. M. V., Bradt, H. V., Dower, R. G. & Jernigan, J. G. *Nature* **269**, 112-116 (1977).
4. Johnson, H. M. *et al. Astrophys. J.* **222**, 664-666 (1978).
5. Thomas, R. M. *et al. Mon. Not. R. astr. Soc.* (in the press).
6. Lewin, W. H. G. *et al. Mon. Not. R. astr. Soc.* **177**, 93P-100P (1977).
7. Buff, J. *et al. Astrophys. J.* **212**, 768-773 (1977).

Interplanetary dust: are there two independent populations?

A RE-EXAMINATION of the information on the size distribution and physical properties of interplanetary dust grains as inferred from space measurements, particularly lunar microcraters, leads to the novel interpretation outlined here in terms of two independent populations. This hypothesis still suffers from uncertainties in existing data but should stimulate new experimental and theoretical work.

New data¹⁻⁵ on lunar microcraters together with the review of Fechtig *et al.*⁶ have firmly established the following features of the cumulative size distribution function $F_c(D_c)$ usually represented by a D_c^{-q} power law: (1) a power index $q = 3.5$ for craters $> 100 \mu\text{m}$; (2) a progressive decrease of q with decreasing diameter, starting at $D_c \approx 100 \mu\text{m}$, and ending with a marked inflection point around $D_c \approx 5 \mu\text{m}$; (3) a new high power index, $q = 3.1$, for microcraters smaller than $\sim 5 \mu\text{m}$ down to $0.1 \mu\text{m}$.

The second point has already been interpreted as evidence for a bimodal population⁷. The third point was recognised only recently^{1,4} when it was realised that accretionary particles and ejecta resulting from larger impacts are extremely efficient in masking the smallest microcraters. Note that microcraters with diameters as small as $0.05 \mu\text{m}$ have been observed⁸. The ratio of crater depth to crater diameter, P/D_c , offers a direct means of probing the density of impacting micrometeorites: although they are still limited to simple experimental conditions, calibrations have shown that P/D_c assumes well defined values characteristic of the density of the projectile^{9,10} (Fig. 1). However, due to conflicting results¹¹, no clear conclusions have been reached on this matter. For instance, Brownlee *et al.*^{7,12} and Mandeville¹³ found a mean density of approximately 3 g cm^{-3} whereas Smith *et al.*¹⁰, Durrani *et al.*¹⁴, and more recently Nagel *et al.*² detected up to three densities: 1, 3 and 8 g cm^{-3} . On the basis of a global study of the published data regrouping 256 microcraters, we propose a clarification of this contradiction as the result apparently depends on the crater sizes (Fig. 1). In particular, by separating them into different size ranges, we found that the apparent contradictory data of Brownlee *et al.*^{7,12} and Smith *et al.*¹⁰ were in satisfactory agreement.

From our analysis (Fig. 1), the following conclusions can be put forward. First, low density grains ($\leq 1 \text{ g cm}^{-3}$) are not detected; second, two families coexist, the large grains having, exclusively, densities typical of silicates and the small ones densities typical of metals with a minor component of the first family. In the intermediate range ($1 < D_c < 10 \mu\text{m}$) the two families are mixed. It is precisely in this range that the cumulative distribution function F_c presents an anomalous behaviour (inflection point) which was an indication for a bimodal population⁷. The *in situ* measurements performed by the Pioneer 8 and 9 and Helios 1 space probes have also conspicuously shown two families of dust grains; the first consists of large grains travelling on nearly circular orbits, whereas the second is composed of small grains, commonly named β -meteoroids¹⁷, leaving the Solar System in hyperbolic orbits^{15,16}. Doppler shifts measurements in the Zodiacal Light Spectrum¹⁸ have confirmed the existence of both types of orbits. The separation between these two families again takes place for particle sizes in the μm range. Collection experiments of extraterrestrial particles performed by Brownlee *et al.*¹⁹ during stratospheric flights have

revealed two distinct kinds of grains: (1) large compact aggregates of submicronic grains having chondritic abundances and a mean density of 2 g cm^{-3} ; (2) smaller, homogeneous, nearly spherical 'iron-sulphur-nickel' (FSN) grains.

The cumulative flux $F(s)$ of interplanetary dust versus grain diameter may be obtained from $F_c(D_c)$, provided the time of exposition and the calibrations of the ratio D_c/s are known. From the work of Morrison and Zinner¹ on sample 12054,54, we used a time of $1.75 \times 10^5 \pm 5 \times 10^4 \text{ yr}$ and a constant value $D_c/s = 2$ from the calibration of Mandeville²⁰ and Bloch *et al.*²¹ to obtain the result presented in Fig. 2.

All the above results are therefore coherent with a two-component model of interplanetary dust: Population 1 consists principally of large grains ($s > 2 \mu\text{m}$) of density typical of silicates or chondritic materials ($2\text{--}3 \text{ g cm}^{-3}$) in nearly circular orbits while Population 2 consists of small grains ($s < 2 \mu\text{m}$) with typically metallic densities (8 g cm^{-3}) in hyperbolic orbits. The dip observed in $F(s)$ (Fig. 2) is readily interpreted as the gradual

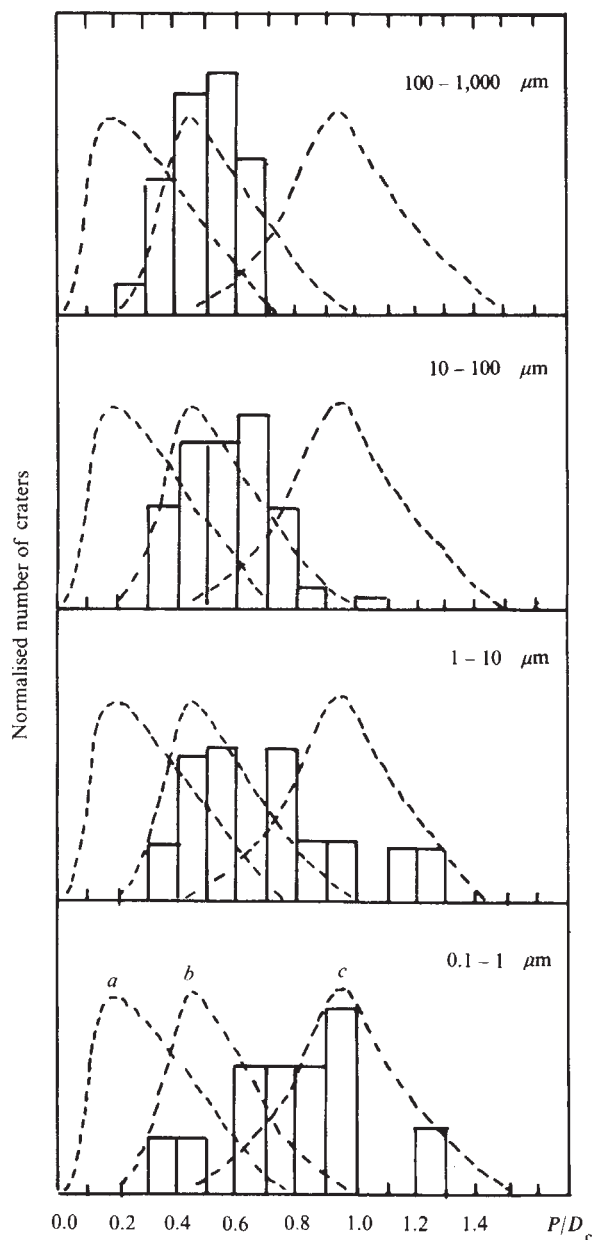


Fig. 1 Normalised histograms of the values of the ratio P/D_c for 256 craters separated in four diameter ranges as indicated. Superimposed (dashed lines) are three simulation curves corresponding to three different densities of impacting projectile. *a*, 1 g cm^{-3} (polystyrene); *b*, 2.7 g cm^{-3} (aluminium); *c*, 7.8 g cm^{-3} (iron).

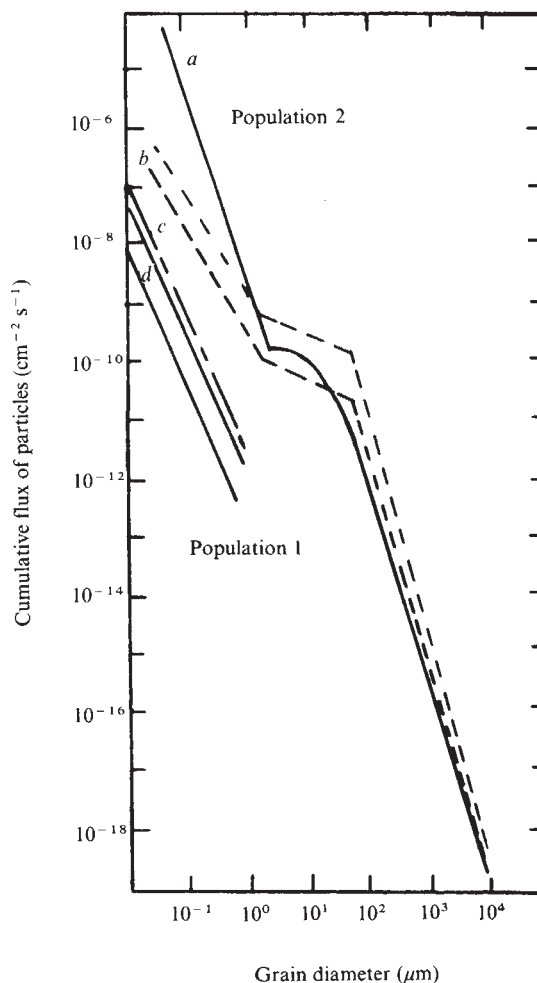


Fig. 2 Cumulative flux plotted against diameter of the particles arriving on the moon: *a*, present result; *b*, results of Fechtig *et al.*⁶; *c*, result derived from the conservation of mass flux; *d*, result from the collisional model²⁶.

disappearance of Population 2 which is more evident on the differential size distribution sketched in Fig. 3.

There have been attempts to explain the origin of β -meteoroids^{17,22} (which we identify with Population 2) based on the collisional fragmentation of larger grains (Population 1) within 1 AU and repulsion of the submicronic fragments under the radiation pressure force. The different physical nature of the two populations deduced from lunar microcraters casts some doubt on this interpretation which may be further rejected on the basis of the conservation of mass flux, an argument which bypasses the difficulties inherent in collisional models of grains (lack of direct calibrations).

Following the model of Zook and Berg¹⁷, Population 1 grains are supposed to spiral slowly towards the Sun under the Poynting-Robertson effect: they may be considered to travel in nearly circular orbits and to collide because of different inclinations. If $f(s) = -dF(s)/ds$ denotes the differential flux of grains measured on the Moon, the spatial density is $f(s) ds / \langle V \rangle$ where $\langle V \rangle$ is the mean velocity of the grains. The inward drift yields a radial velocity²³ $V_{PR}(s)$ leading to a radial flux of grains:

$$V_{PR}(s)f(s) ds / \langle V \rangle$$

The total mass influx at 1 AU is therefore

$$\Phi_1 = \frac{\pi}{6} \frac{\rho}{V} \int_{s_{1,\min}}^{s_{1,\infty}} s^3 V_{PR}(s) f(s) ds$$

where ρ is the density of the grains, $s_{1,\min} \approx 2 \mu\text{m}$ is the minimum diameter, and $s_{1,\infty}$ is the maximum diameter of Population 1 grains.

The situation is entirely different for Population 2 grains for which the radial velocity is of the order of the orbital hyperbolic velocity. The total mass outflux at 1 AU is simply

$$\Phi_2 = \frac{\pi}{6} \rho \int_{s_{2,\min}}^{s_{2,\infty}} s^3 f(s) ds$$

where $s_{2,\min} \approx 0.01 \mu\text{m}$ is taken as the minimum size of Population 2 grains and $s_{2,\infty} \approx 2 \mu\text{m}$ as the maximum size. The final value of Φ_2 is not sensitive to this choice of parameters. The collisional hypothesis requires

$$\Phi_1 \geq \Phi_2$$

while we found

$$\Phi_1 \approx 5 \times 10^{-23} \text{ g cm}^{-2} \text{ s}^{-1} \quad \text{and} \quad \Phi_2 \approx 5 \times 10^{-20} \text{ g cm}^{-2} \text{ s}^{-1}$$

The difference between the two fluxes is so large that our present conclusion will not be affected by any possible variations of the parameters involved in the calculation (grain density, orbital velocities) not by uncertainties in the exposure time which do not affect the ratio Φ_1/Φ_2 . Using the above value of Φ_1 , we can give an upper limit for the flux of fragments resulting from collisional destruction under the constraint $\Phi_2 = \Phi_1$ assuming further that (1) all grains of Population 1 are destroyed and all fragments smaller than $2 \mu\text{m}$ are ejected on hyperbolic orbits; (2) the distribution of fragments follows a power law $s^{-\eta}$ where $\eta = 3.48$ (refs 24, 25). The result is shown in Fig. 2 and is three orders of magnitude smaller than the observed flux. We constructed an elaborated collisional model²⁶ using the basic procedure of Dohnanyi²² but incorporating as many real data as possible^{19,25}. The model fluxes of fragments always lie slightly under the upper limit deduced from the conservation of mass flux. If one accepts the present results from lunar samples and space experiments, we arrive at the conclusion that the two populations are entirely independent on several grounds: size, density and dynamics.

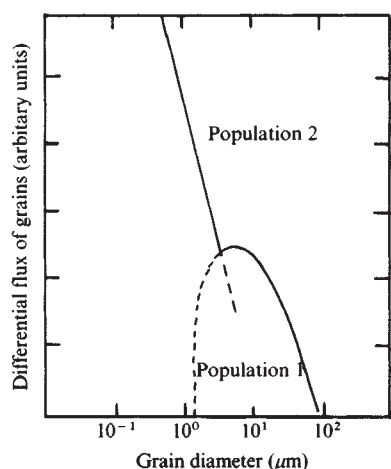


Fig. 3 Differential flux of particles as a function of particle diameter (arbitrary units).

Population 1 has probably its origin in dust grains ejected by short period comets. The sizes and size distribution law are in surprisingly good agreement with those derived for Comet d'Arrest²⁷. The bulk of Population 1 grains probably disappear within 1 AU as the value of β (ratio of the radiation pressure to the gravitational forces) for silicates²⁸ and the 'initial' orbital conditions at 1 AU (nearly circular orbits) ensure that they will not escape the inner region of the Solar System whatever processes (collisions, sublimation) they undergo. We may therefore consider that Φ_1 corresponds to the mass loss rate of Population 1. Integrating Φ_1 taking into account the distribution of inclination²², we found a mass loss rate of $2.8 \times 10^4 \text{ g s}^{-1}$. To

maintain Population 1 in steady state, approximately 70 short-period comets similar to Comet d'Arrest is required on the basis of an average production rate of 400 g s^{-1} . Note that this satisfactory number is obtained assuming a grain albedo of 0.5, a more reasonable value than the 0.001 considered by Sekanina and Schuster²⁷. Population 2 presents far more difficult problems. Either the grains are formed in the solar atmosphere and subsequently expelled by radiation pressure²⁹ (although this presents difficulties³⁰) or they are still contained in larger bodies when they penetrate inside 1 AU. Long-period comets could be a possible source as they are known to be very dust productive. However, the existence of only prograde hyperbolic orbits¹⁸ seems to rule out this solution to a large extent.

L. B. LE SERGEANT
P. L. LAMY

Laboratoire d'Astronomie Spatiale,
Les Trois Lucs,
13012 Marseille, France

Received 4 September; accepted 8 November 1978.

- Morrison, D. A. & Zinner, E. *Proc. Lunar Sci. Conf. VIII*, 841, Houston (1977).
- Nagel, K., Neukum, G., Dohnanyi, J. S., Fechtig, H. & Gentner, W. *Proc. Lunar Sci. Conf. VII*, 1021, Houston (1976).
- Mandeville, J. C. *Proc. Lunar Sci. Conf. VII*, 1031, Houston (1976).
- Mandeville, J. C. *Proc. Lunar Sci. Conf. VIII*, 883, Houston (1977).
- MacDonvw J. A. M. *Proc. Lunar Sci. Conf. VIII*, 643, Houston (1977).
- Fechtig, J. *et al. Proc. Soviet-American Conf. Cosmochemistry of the Moon and the Planets*, Moscow (1975).
- Brownlee, D. E., Hörz, F., Hartung, J. B. & Gault, D. E. *Proc. Lunar Sci. Conf. VI*, 3409, Houston (1975).
- Poupeau, G., Walker, R. M., Zinner, E. & Morrison, D. A. *Proc. Lunar Sci. Conf. VI*, 3433, Houston (1975).
- Vedder, J. F. & Mandeville, J. C. *J. geophys. Res.* **79**, 3247 (1974).
- Smith, D., Adams, N. G. & Khan, H. A. *Nature* **252**, 101 (1974).
- Hartung, J. B. in *Lect. Not Phys.* **48**, 209 (1976).
- Brownlee, D. E., Hörz, F., Vedder, J. F., Gault, D. E. & Hartung, J. B. *Proc. Lunar Sci. Conf. IV*, 3197, Houston (1973).
- Mandeville, J. C. *Proc. Lunar Sci. Conf. VI*, 3403, Houston (1975).
- Durrani, S. A., Khan, H. A., Bull, R. K., Dorling, C. W. & Fremlin, J. H. *Proc. Lunar Sci. Conf. V*, 2453, Houston (1974).
- Berg, O. E. & Grün, E. *Space Res.* **13**, 1047 (1973).
- Grün, E., Fechtig, H., Kissel, J. & Gamelin, J. *Geophys. J.* **42**, 6 (1977).
- Zook, H. A. & Berg, O. E. *Planet. Space Sci.* **23**, 183 (1974).
- Fried, J. W. *Astr. Astrophys.* **68**, 259 (1978).
- Brownlee, D. E., Hörz, F., Tomandl, D. A. & Hodge, P. W. *IAU Colloqu. no. 25*, NASA SP-393, 962 (1976).
- Mandeville, J. C. thesis, Univ. Paul Sabatier, Toulouse (1972).
- Bloch, M. R., Fechtig, H., Gentner, W., Neukum, G. & Schneider, E. *Proc. Lunar Sci. Conf. II*, 2639, Houston (1971).
- Dohnanyi, J. S. in *Lect. Not Phys.* **48**, 170 (1976).
- Wyatt, S. P. & Whipple, F. L. *Astrophys. J.* **111**, 134 (1950).
- Gault, D. E. & Heitowit, E. D. *Proc. 6th Hypervelocity Impact Symp.* **2**, 419 (1963).
- Fujiwara, A., Kamimoto, G. & Tsukamoto, A. *Icarus* **31**, 277 (1977).
- Le Sergeant, L. & Lamy, Ph. (in preparation).
- Sekanina, Z. & Schuster, H. E. *Astr. Astrophys.* **65**, 29 (1978).
- Lamy, Ph. *Astr. Astrophys.* **35**, 197 (1974).
- Hememway, C. L. in *Lect. Not Phys.* **48**, 251 (1976).
- Mullan, D. J. *Astr. Astrophys.* **61**, 369 (1977).

Is the Sun almost-intransitive?

THERE is now ample instrumental, historical and proxy evidence for palaeoclimatic variability over time scales ranging from 10^3 to 10^9 yr (ref. 1). Although a great deal of research is still needed to establish more precisely the nature and finer spatial and temporal resolution of these changes, the main theoretical problem is to understand the mechanisms responsible for this variability over such a wide range of time scales. We suggest here that if the Sun were almost-intransitive, it could explain some of this climatic variability. Many hypotheses have been advanced by researchers in various disciplines based on physical processes, both terrestrial and extra-terrestrial¹⁻⁸, and although there is no general consensus as to the causes of variability over various time scales, the common view among climatologists is that the very long time-scale fluctuations with periods of $\sim 10^8$ yr can be explained by causes of terrestrial origin, such as the changes in oceanic and continental geometries^{3,10}. The intermediate time-scale fluctuations with

periods of $\sim 10^4$ – 10^5 yr are thought to have been caused by variations in Earth's orbital elements⁹. The relatively short time-scale fluctuations with periods of $\sim 10^2$ – 10^4 yr are, on the other hand, hard to explain, because of the clear lack of a physical process with the appropriate time scale. Solar variability is, however, accepted as a serious contender^{3,10}. All these hypotheses, whether terrestrial or extra-terrestrial, have in common the assumption that the changes in climate are deterministic, that is, they are a causal response of the climatic system to a change in some environmental (internal or external) parameter, even though the actual mechanism of change is not usually understood.

A radically different hypothesis by Lorenz, suggests that the climatic variability over some time scales may turn out to be non-deterministic^{14–16}. This possibility arises, essentially, from the fact that the climate can be defined as a set of long-term statistics of the time-dependent solution of a set of appropriate non-linear differential equations. On the basis of ergodic theory, both transitive systems (ones in which long-term statistics of some time-dependent solution is independent of the choice of boundary conditions) and intransitive systems (ones in which the solution is dependent on the choice of the boundary conditions) can exist¹⁶. Evidence for intransitive systems also exists both experimentally¹⁷ and numerically¹⁴. In addition to the above two modes of behaviour for a system, Lorenz has pointed out the possibility that a system of non-linear differential equations may have solutions which exhibit different statistical properties within different segments of a long time span, with a completely constant environment. Such systems he has termed almost-intransitive^{14–16}. As an example of these systems, he has constructed numerical models (which to some extent resemble the numerical models of the Earth's atmosphere, or for that matter any similar non-linear system of differential equations), which behave in an almost-intransitive fashion¹⁶.

There is also some evidence to suggest that the climate may have behaved in an almost-intransitive fashion over time scales with periods of $\sim 10^4$ yr (ref. 18). It may, therefore, seem likely that the possible almost-intransitivity of the climatic system could account for climatic variability over such time scales, without the need to invoke solar variability. This, however, does not seem to be the case, as there are strong indications of solar variability over these time scales, both historical¹² and proxy. The latter evidence is based on the measurements of radiocarbon (^{14}C) fluctuations in tree rings which are thought to have been caused by the solar modulation of galactic cosmic rays¹¹. There is also some apparent correlation between the radiocarbon variations and the historical climatic record^{10,11,19,20}. Solar variability over such time scales cannot, therefore, be ruled out in favour of almost-intransitivity of the climatic system.

There is, however, some difficulty in accounting for the solar variability over these time scales. This becomes apparent when one takes a closer look at the commonly accepted solar model (with its own difficulties²¹), and the time scales characteristic of the physical processes that are thought to be operative in the Sun. Such examination reveals a spectrum of time scales dominated on one end with very long time scales of $\sim 10^7$ – 10^{13} yr, and on the other with very short time scales of $\sim 10^{-4}$ – 10^{-1} yr (ref. 13). Therefore within the context of the presently accepted solar model, the time scales of variability for which there is some evidence and are thought to be important climatologically, namely those with periods of $\sim 10^2$ – 10^4 yr, are hard to explain. It therefore seems to be necessary to reconcile the evidence for solar variability over such periods, the lack of physical processes with appropriate time scales operative in the Sun and the possibility of the existence of systems which behave almost-intransitively.

We suggest that this could be achieved if the Sun itself (its coupled 'convective zone-atmosphere') were almost-intransitive. Some of the arguments that support the plausibility of this hypothesis and some of its consequences are given below.

The complete, unapproximated fluid dynamical equations representing the convective zone-atmosphere of the Sun are

highly non-linear and are very similar to those representing the Earth's atmosphere. Lorenz's arguments regarding the plausibility of the almost-intransitivity of the climatic system could therefore equally apply to the Sun, or to any physical system where similar highly non-linear differential equations are operative.

Almost-intransitivity of the Sun could account for the evidence of solar variability in radiocarbon data and in the historical record, and also the climatic variability over these time scales. Almost-intransitivity of the climatic system on the other hand, although possible, cannot account for the independent evidence for solar variability over such time scales.

There is also the possibility, on the basis of dishpan experiments¹⁷ with their limited similarity to the climatic system, that the climatic system may be intransitive¹⁶. If so, it is hard to switch from one climatic regime to another without the presence of an external forcing mechanism¹⁶. Almost-intransitivity of the Sun could act as such an external forcing agency.

It seems plausible therefore that the Sun could be almost-intransitive over some time scales. It should, however, be added that the climatic system is highly non-linear and complicated, and it is possible that even if the Sun is almost-intransitive, this need not be the sole cause of climatic variability over such time scales. Nevertheless, if the Sun turns out to be almost-intransitive it could have astronomical as well as climatological significance.

I thank my colleagues at the Climatic Research Unit and at Queen Mary College, London for their encouragement, and the US Office of Naval Research for grant No. 0014-77-9-0074.

R. K. TAVAKOL

*Climatic Research Unit,
School of Environmental Sciences,
University of East Anglia,
Norwich, UK*

Received 14 September; accepted 10 November 1978.

1. *Understanding climatic change: a program for action*, Report of GARP Panel on climatic variations (US National Academy of Sciences, Washington, D.C. 1975).
2. Mitchell, J. M., Jr *Meteor. Monogr.* **8**, 155 (1968).
3. Mitchell, J. M., Jr *Quat. Res.* **6**, 481 (1976).
4. Kutzbach, J. E. *Quat. Res.* **6**, 471 (1976).
5. McCrea, W. H. *Nature* **255**, 607 (1975).
6. Clark, D. H., McCrea, W. H. & Stephenson, F. R. *Nature* **265**, 318 (1977).
7. Wdowczyk, J. & Wolfendale, A. W. *Nature* **268**, 510 (1977).
8. Begleman, M. C. & Rees, M. J. *Nature* **261**, 229 (1976).
9. Hays, J. D., Imbre, J. & Shackleton, N. J. *Science* **194**, 1121 (1976).
10. Hays, J. D. in *The Solar Variability and its Variations*, 73 (ed. White, O. R.) (Colorado Associated Press, Boulder, 1976).
11. Damon, P. E. in *The Solar Variability and its Variations*, 429 (ed. White, O. R.) (Colorado Associated Press, Boulder, 1976).
12. Eddy, J. A. in *The Solar Variability and its Variations*, 51 (ed. White, O. R.) (Colorado Associated Press, Boulder, 1976).
13. Gough, D. O. in *The Solar Variability and its Variations*, 451 (ed. White, O. R.) (Colorado Associated Press, Boulder, 1976).
14. Lorenz, E. N. *Meteor. Monogr.* **8**, 1 (1968).
15. Lorenz, E. N. *J. appl. Meteor.* **9**, 325 (1970).
16. Lorenz, E. N. *Quat. Res.* **6**, 495 (1976).
17. Hide, R. *Phil. Trans. R. Soc. A250*, 441 (1958).
18. Flohn, H. *GARP Publications Series*, No. 16 (WMO-ICSU, Geneva, 1975).
19. Suess, H. *Meteor. Monogr.* **8**, 146 (1968).
20. Damon, P. E. *Meteor. Monogr.* **8**, 151 (1968).
21. Roxburgh, I. W. *Proc. 2nd European Conf. on Solar Physics* (ed. Rosch, E. in the press).

Waves in narrow channels: faster capillary waves

DURING experiments on the propagation of water waves in channels with parallel sides made of Perspex, it was found that the phase velocities of small-amplitude waves with wavelengths in the range 40–100 mm were consistently higher than predicted by the standard hydrodynamic theory^{1,2}. The original observations of this anomaly were made in a channel 100 mm wide filled to a depth of 50 mm. For these none too small dimensions, the standard theory might be expected to provide a close approximation to the wave velocity, but velocities $\sim 0.5\%$ too high were observed. Highly surface-clean water and clean apparatus was used in all the experiments, and with these

precautions the measurements proved to be wholly reproducible, prompting a closer examination, which is described here, of the wave motion to detect the cause of the anomaly. It was seen that the lines of contact between the water surface and the sides of the channel did not in fact rise and fall in step with the movement of the surface at the centre of the channel, instead remaining fixed in their original locations with the water at rest.

The observed immobility of the contact lines exemplifies the phenomenon commonly called contact-angle hysteresis. A larger contact angle is apparent when a liquid is advancing over a solid surface than when it is receding; and correspondingly, for a static contact line, there is an intermediate range over which the angle can vary to counterbalance forces tending to dislodge the line. This hysteresis, which often spans a sizable fraction of the mean contact angle, may be observed easily with almost every practical (non-reacting) combination of liquid and solid, and it is usually due to either microscopic roughness or chemical inhomogeneity of the solid surface.

With regard to experiments such as ours, therefore, it is reasonable to conclude that contact lines will remain fixed, provided the wave amplitude does not exceed a certain threshold. The vibrating surface of the water will then curve like a membrane tied at its edges, and the capillary pressures associated with the additional, transverse curvature will evidently stiffen the surface, so increasing the velocity of wave propagation. Further experiments were made using much narrower channels, for which the stiffening effect was expected to be correspondingly greater, and phase velocities were recorded that were remarkably higher than those of the more familiar, long-crested waves in the same range of wavelengths. For example, in a channel 4 mm wide filled to a depth of 25 mm, the phase velocity of waves 100 mm long was found to be about 2.5 times the predicted value for waves of the same length free from edge constraints.

A linearised theory for the new class of wave motions has been developed, focussing on the case of an exactly brimful channel. An advantage of this case as a theoretical model is that the free surface is flat when at rest, so that evaluation of the capillary pressures on the perturbed surface is comparatively simple. It also has considerable practical advantages, in particular that the contact lines can be located securely along sharp edges of the channel and that the required flatness of the free surface can be judged accurately by viewing the mirrored images of distant objects; accordingly, it was made the subject of the final, most exacting series of experiments devoted to these waves. Other assumptions of the theory apparently compatible with the experimental conditions are that the liquid is incompressible and inviscid, so that its motion started from rest is irrotational, and that the liquid has a constant surface tension.

The theory is more complex than the corresponding theory for waves free from edge constraints; and even in the simplest case where the channel has a rectangular cross-section, explicit solutions are not available. As usual in water-wave theory, the hydrodynamic loading of the perturbed free surface can be represented abstractly by a Neumann boundary-value problem, but the edge conditions imposed on functions representing the displacement of the free surface imply that no finite combination of the eigenvectors of this constituent problem can solve the complete problem including the effects of surface tension. A variational formulation leads, however, to a well-defined expression, in the form of a Rayleigh quotient, for the possible frequencies of waves with prescribed wavelength, and this result served both to establish that the complete problem has solutions in appropriate function classes and to give useful estimates of wave properties.

It seems that the system, like the simpler one without edge constraints, has a fundamental wave mode which can propagate at any given frequency, and which for a given wavelength has the lowest possible frequency. Higher modes have respective cut-off frequencies, which are the frequencies of standing waves independent of distance along the channel, and below which

propagation is impossible. For each of the higher modes, the phase velocity is unbounded and the group velocity is zero in the limit as frequency approaches the cutoff value from above. As might be expected, the fundamental mode uniquely has the property that the edges and perpendiculars to them are its only nodal lines: in the cross-section through a longitudinal wave crest, the elevation of the free surface above its undisturbed level is symmetrical about a maximum at the centre, vanishing only at the edges. A less obvious fact emerging from the theory is that at the edges the curvature of the free surface is generally not zero, vanishing only where there is a spanwise nodal line.

For comparison with the experimental results, the theoretical relationship between frequency and wavelength was estimated by means of the Rayleigh–Ritz method, which is particularly amenable in the relevant case of rectangular channels. Taken to two stages of approximation, the method gave results for the fundamental mode in very satisfactory agreement with experiment, and such is the efficiency of the method that even the roughest, first-order estimates seemed to be close to the true values. For example, applied to channels whose breadth, b , is less than ~ 20 mm and whose depth, h , is not less than b , the following formula gives a good estimate of the fundamental phase velocity, c_1 , as a function of wave-number α ($= 2\pi/\text{wavelength}$):

$$c_1^2(\alpha) = \frac{1.2(g + \gamma\alpha^2) + 12\gamma b^{-2}}{\alpha\{\coth(\alpha h) + 0.03050(\alpha b) - 0.000376(\alpha b)^3\}}$$

Here g is the gravity constant, and γ is the ratio of surface tension to density. This formula is obtainable simply by taking the spanwise distribution of the oscillatory surface displacement to be proportional to $f(x) = (x/b) - (x/b)^2$, where x is the distance from one edge, substituting the function f in the Rayleigh quotient whose absolute minimum expresses $\omega_1^2 = \alpha^2 c_1^2$, and introducing certain elementary inequalities satisfied by the infinite series whose sum is the denominator of the quotient. The theory proved the formula to be a strict upper bound for $c_1^2(\alpha)$, and it should be particularly close to the true value when αb is much smaller than 1. Over the range of our experiments, the values of $c_1(\alpha)$ predicted by it were at most about 3% in excess of the measured values. Note that in the long-wave limit $\alpha \rightarrow 0$, the estimate is $c_1^2(0) = (1.2g + 12\gamma b^{-2})h$, whereas the corresponding well-known result for waves without edge constraints is $c^2(0) = gh$.

Figure 1 presents values of wavelength measured in two channels with different rectangular cross-sections, and also includes theoretical curves for the fundamental mode respective to the two sets of experimental conditions. The Perspex channels, each having a working length of about 1 m and an additional length of about 450 mm with a beach to absorb the waves, were filled exactly to their brims which had been machined smooth. Waves were excited at constant frequency by a small oscillating paddle (made of PTFE) which was driven by an electrodynamic vibrator, and the range of frequencies employed was such that only waves in the fundamental mode would propagate away from the wave-generator. Relative to the original electronic signal driving the motion, variations in the phase of the waves with distance along the channel were monitored by means of a capacitance probe which could be continuously positioned at a fixed height above the water surface, and wavelengths were measured by observing the distances between positions of coincident phase, indicated by Lissajous figures on an oscilloscope.

On the whole the new theoretical results are in good agreement with the experimental observations, although the measured phase velocities tend to be slightly less than those predicted. This discrepancy is in the right direction for it to be accountable in part to the residual error implicit in the use of the Rayleigh–Ritz method, but another possible reason for it can be linked to the difficulty of ensuring a perfectly clean water surface

in the channels. Their narrowness made them more or less inaccessible for careful cleaning except by high-speed water jets, and measurements of surface tension *in situ* were impracticable. The water used in the experiments was unable to sustain surface bubbles after being shaken, thus passing a sensitive test for the absence of surface-active contamination, and it was put into the channels by a glass pipette which had been cleaned with chromic acid. As used in the calculations, the value of surface tension was taken to be that for the supply of surface-clean water.

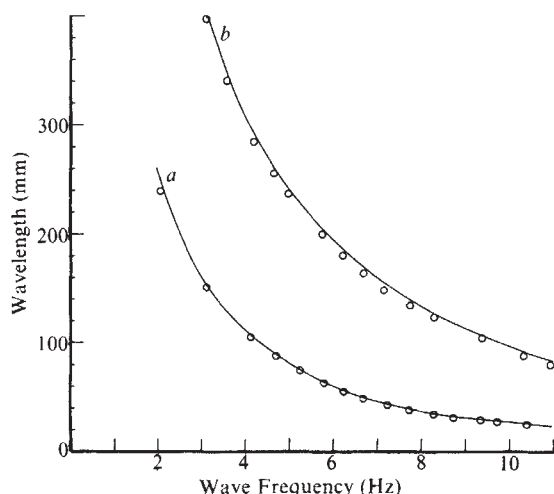


Fig. 1 Measured values of wavelength as a function of frequency, for water waves of small amplitude in two channels: *a*, width 20.3 mm, depth 21.5 mm; *b*, width 4.0 mm, depth 25.0 mm. The solid curves were calculated by the Rayleigh-Ritz method taken to two stages of approximation. Surface tension was estimated to be 72.0 mN m^{-1} .

Contamination subsequently picked up, however, would have had effects consistent with what was actually observed. The discrepancy between measured and predicted wave properties did in fact tend to increase gradually during experimental runs, consistently with a gradual decrease of surface tension.

Capillary wave phenomena depending on edge conditions of this kind seem to have attracted little previous attention and many problems akin to the present one may be suggested for theoretical as well as experimental study. The most promising candidate is perhaps the case of standing waves in closed basins, whether brimful or with fixed contact lines supporting upwards or downwards curving menisci. The nonlinear theory of progressive waves with moderately large amplitudes also poses certain challenging problems. A fuller account of our investigation including a rigorous treatment of the theoretical problem posed by the new class of water waves will be published elsewhere³.

We thank the NERC for support, also J. K. Bartington and J. E. Davis for technical assistance.

JOHN C. SCOTT
T. BROOKE BENJAMIN

Fluid Mechanics Research Institute,
University of Essex,
Wivenhoe Park,
Colchester, Essex, UK

Received 7 September; accepted 3 November 1978.

1. Lamb, H. *Hydrodynamics*, 6th edn, sect. 267 (1932).
2. Ramsey, A. S. *A Treatise on Hydromechanics*, 4th edn, sect. 10.61 (1935).
3. Benjamin, T. B. & Scott, J. C. *J. Fluid Mech.* (in the press).

SSTs, nitrogen fertiliser and stratospheric ozone

THE threat to stratospheric ozone from supersonic transport (SST) exhaust emissions has been a subject of controversy since Hampson¹ first suggested that significant ozone reductions might be caused by discharged water vapour. Later, after Johnston² and Crutzen³ pointed out that even larger ozone depletions might result from emitted nitrogen oxides, water vapour effects were considered to be relatively unimportant. In the intervening years, predictions of SST ozone modification have evolved with advancing scientific knowledge⁴⁻⁷. Using a recently revised model of the stratosphere⁸, we report here that a substantial ozone layer enhancement could accompany worldwide SST fleet operations, and that water vapour may be an important factor in SST assessments. We have also found that increased nitrogen fertiliser use might, likewise, enhance the ozone layer, in contrast with previous calculations of ozone reduction⁹⁻¹¹.

Zahniser and Howard¹² have just reported the first direct measurement of the rate coefficient for the reaction

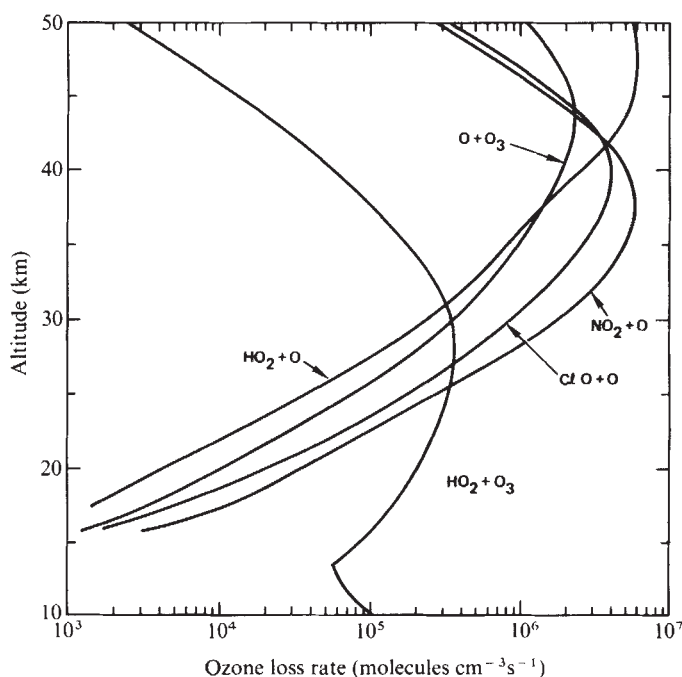


Their result, $(1.4 \pm 0.4) \times 10^{-14} e^{-(580 \pm 100)/T} \text{ cm}^3 \text{ s}^{-1}$ is, at stratospheric temperatures, a factor of 4-6 larger than earlier values, which were based on reaction rate ratios¹³. In the lower stratosphere, where most terrestrial ozone resides, reaction (1) accelerates ozone destruction through a catalytic chemical cycle that also involves the reaction



Figure 1 gives the rates of chemical destruction of odd-oxygen (essentially ozone) in the unperturbed stratosphere due to the dominant catalytic reaction cycles involving hydrogen radicals ($\text{HO}_x \equiv \text{OH}$ and HO_2), nitrogen oxides ($\text{NO}_x \equiv \text{NO}$ and NO_2), free chlorine gas ($\text{Cl}_x \equiv \text{Cl}$ and ClO), and the Chapman reactions among the oxygen allotropes. The few chemical processes shown in Fig. 1 represent the most important ozone loss mechanisms in current models of the stratosphere; they are

Fig. 1 Ambient rates of consumption of stratospheric odd-oxygen (ozone) for key catalytic reactions steps.



responsible for more than 80% of the total ozone loss rate. A complete description of the one-dimensional model used here to make ozone calculations can be found elsewhere¹⁴.

Note in Fig. 1 that HO_x , because of reaction (1), is now the major catalytic agent for ozone loss below 25 km; HO_x also dominates ozone loss above 40 km. Between 25 and 40 km, NO_x has the greatest influence on the ozone loss rate. The Cl_x and Chapman reactions represent important secondary odd-oxygen loss mechanisms (assuming a current atmospheric chlorine level of about 2 parts per billion by volume). The stratospheric ozone content is roughly determined by the balance between loss (as in Fig. 1) and production through O_2 photolysis. The implications for ambient ozone of the new rate constant for reaction (1) are discussed by Whitten *et al.*¹⁵; some of these implications had been speculated on in earlier publications^{6,7}.

Shortly before the completion of the Zahniser and Howard¹² study, another major change in aeronomic models occurred when Howard and Evenson¹⁶, and Burrows *et al.*¹⁷ redetermined the rate constant for



The revised rate coefficient, $\sim 8 \times 10^{-12} \text{ cm}^3 \text{ s}^{-1}$ at room temperature, is nearly a factor of 40 greater than previously adopted stratospheric values. The aeronomic effects of reaction (3) are complex^{7,8}. Basically, the larger rate constant increases the importance of HO_x relative to NO_x as an ozone-active catalyst, which amplifies the role of reaction (1) in the ozone balance.

Supersonic aircraft cruise at heights well into the stratosphere and release NO_x and water vapour as exhaust products. In a recent detailed assessment of the potential environmental impact of future SSTs, Poppoff *et al.*⁸ showed that NO released by SSTs inhibits the normal HO_x catalysis of ozone because reaction (3) diverts HO_2 from reaction (1). Added NO_x also induces faster recombination of HO_x radicals through a reaction of OH with nitric acid, although NO_x regenerates some HO_x (and ozone) through a natural smog mechanism involving methane oxidation⁸. Moreover, NO—which has a fast reaction

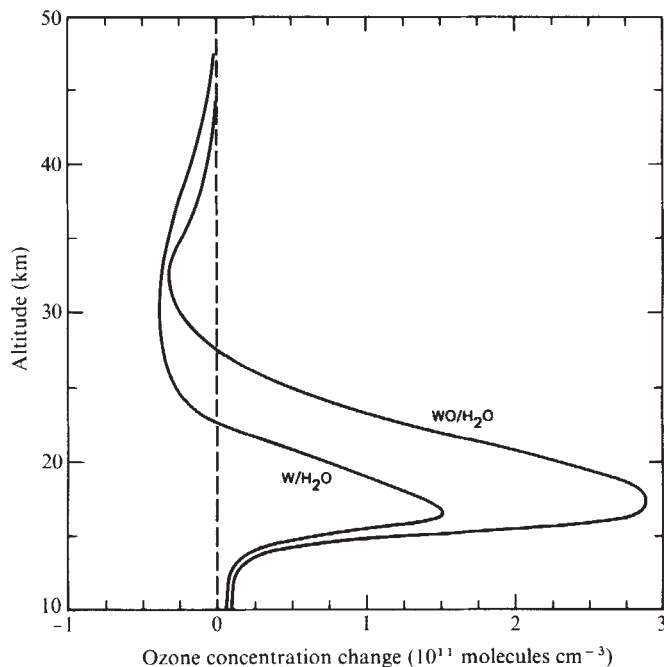


Fig. 3 Calculated steady-state stratospheric ozone concentration changes due to an NO_x injection of $7 \times 10^8 \text{ kg NO}_2 \text{ yr}^{-1}$ (globally averaged) both with and without water vapour emission.

with ClO analogous to reaction (3)—interferes with chlorine-sensitized ozone destruction. As a result of these coupled interactions, the total content of stratospheric ozone was calculated to increase slightly (a few tenths of a per cent) for modest-sized fleets of SSTs (up to several hundred planes) flying below about 25 km.

However, Poppoff *et al.*⁸ also found that aircraft water vapour emissions may cause significant chemical degradation of ozone because of the following circumstances: (1) H_2O is the major source of stratospheric hydrogen radicals, which now dominate ozone loss throughout most of the stratosphere (see Fig. 1); (2) advanced aircraft technology can (theoretically) suppress NO_x emissions, but H_2O emissions are essentially fixed in proportion to the fuel burned. It has been proposed that ozone depletions due to stratospheric water vapour increases might be largely offset by ozone recovery resulting from H_2O infrared cooling of the air, which slows some of the ozone catalytic reactions¹⁸. Unfortunately, definitive thermal-feedback calculations related to SST perturbations have yet to be made.

Figure 2 gives our revised estimates of ozone column changes resulting from NO_x injection at 20 km. Two injection cases are considered: (1) NO_x injection only, and (2) simultaneous NO_x and H_2O injection assuming relative engine emission indices of 6 g NO_2 per kg fuel (a design goal), and 1.3 kg H_2O per kg fuel (stoichiometrically determined for jet fuel combustion). Both cases neglect possible thermal feedback effects. Case 2 roughly corresponds to 550 H_2O molecules released with each NO molecule (where NO is taken to be the predominant NO_x species emitted by aircraft engines).

Most of the present calculations were made with the one-dimensional model mentioned earlier using a Wofsy-McElroy type (WT) of eddy diffusion coefficient¹⁹ with the tropopause height adjusted to 13 km (ref. 8); a few cases using a Dickinson-Chang type²⁰ (DC) are shown for comparison. Note in Fig. 2 that, with regard to the major effects of SSTs on ozone, the differences between diffusion coefficients are not significant. It should also be pointed out that the uncertainty in such model predictions is large^{8,20}.

According to our latest estimates (Fig. 2), NO_x injection alone could lead to a surprisingly large total ozone enhancement—several per cent or more. In contrast, the ozone content above

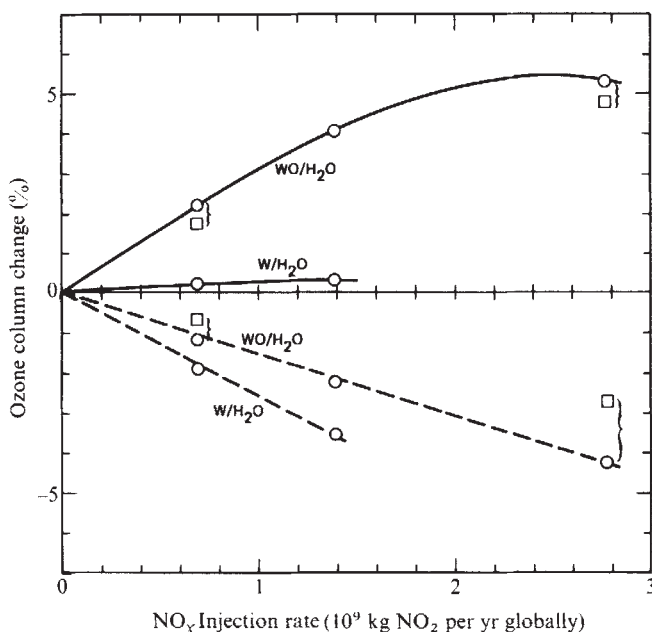


Fig. 2 Steady-state perturbations of the vertical ozone column above 10 km (solid line) and 30 km (dashed line) as a function of the rate of a steady NO_x injection at 20 km. Results are shown both with ($\text{W}/\text{H}_2\text{O}$) and without ($\text{WO}/\text{H}_2\text{O}$) simultaneous water vapour release. O, Values computed using a Wofsy-type (WT) diffusion coefficient; □, Values computed with an alternative (DC) diffusion profile (see text).

30 km would be significantly depleted. Substantial total ozone enhancements are predicted even for the largest NO_x injection rate considered, which corresponds to more than 1,000 advanced aircraft operating on a daily worldwide basis. The reason for these large ozone enhancements can be traced directly to the increased rate constants for reactions (1) and (3). That is, rapid HO_x catalysis of ozone through reaction (1) is strongly suppressed by injected NO through reaction (3), more so in our revised model than in the model used by Poppoff *et al.*⁸.

When NO_x and H₂O are injected simultaneously, we still find ozone increases, albeit much smaller ones. Nonetheless, the magnitude of the chemical degradation of ozone by added water vapour is a factor of 2 larger in our updated model because of the faster rate for reaction (1). Calculated ozone changes due to added NO_x and H₂O are quite sensitive to the adopted background chlorine level (2 parts per billion by volume here). Generally, as the chlorine level rises, the ozone enhancement following NO_x injection, and the ozone decrease following H₂O injection, both become larger⁸.

Figure 3 shows the predicted ozone concentration changes corresponding to the operation of about 200 aircraft at 20 km. Water vapour chemical reactions are seen to negate in large part the NO_x-induced ozone gains calculated below 25 km. The possible significance to atmospheric temperature and dynamics of such ozone alterations has never been treated satisfactorily.

In the past, the related problem of nitrogen fertiliser and sewage decomposition into free nitrous oxide (N₂O) with subsequent photochemical (NO_x) effects on ozone has been stated in terms of the per cent ozone depletion per unit increase in atmospheric N₂O (refs. 9–11). Using our revised model, however, we calculate that a 50% increase in N₂O abundances could lead to a 2.7% increase in total ozone (above 10 km, for the WT diffusivity). Even so, the ozone content above 30 km would be reduced by more than 5%.

In accordance with our latest estimates of global ozone modification due to stratospheric flight, new research emphasis should be placed on the aeronautical role of water vapour effluence at high altitude, with special regard to its photochemical impact on ozone and infrared effect on air temperature.

R & D Associates,
Marina del Rey, California 90291

NASA Ames Research Center,
Moffett Field, California 94035

Department of Meteorology,
San Jose State University,
San Jose, California 95192

R. P. TURCO

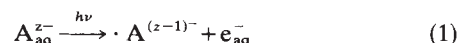
R. C. WHITTEN
I. G. POPPOFF

L. A. CAPONE

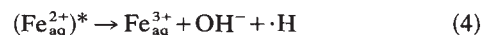
Precambrian solution photochemistry, inverse segregation, and banded iron formations

SOLAR radiation on an early Precambrian sea would generate short-lived excited species near the surface of the water. Many of the species would be powerful oxidising or reducing agents. Despite back-reactions, net reducing power would be lost to the atmosphere. This effect would be the opposite of present day photosynthesis and can be termed 'inverse segregation'. Its possible relevance to banded iron formation is discussed here.

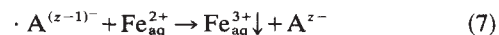
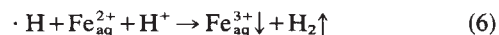
Consider an early Precambrian sea on a summer's day. The atmosphere is neutral (N₂+CO₂) or slightly reducing (+also a little H₂)¹. High energy UV sunlight (from ~200 nm) thus reaches the seawater which contains Fe²⁺ among other simple inorganic ions. Many of these absorb radiation in the 200–300 nm range, often through charge-transfer-to-solvent transitions^{2–6}. As a result numerous short-lived excited species are generated in the upper few centimetres of the water, for example:



Many of these are powerful oxidising and reducing agents and may generate others through, for example:



Exceedingly complex cross reactions and back reactions follow, but not very far below the surface only longer lived oxidised and reduced species are found. These continue to react, often with each other. They cannot fully neutralise each other, however, because during these processes some reduced species have escaped to the atmosphere. In particular H₂, but probably also CO and some of the small organic molecules formed following reactions such as (5) (ref. 6). By contrast oxidising species, particularly radicals⁴, have quickly converted Fe²⁺ to Fe³⁺ which precipitates as highly insoluble ferric and ferric-ferrous hydrolysis products. For example:



There are, perhaps, still back reactions with reduced species remaining in deeper waters, but net oxidising power remains in the sediments in balance with the reducing power lost to the atmosphere. This effect would be the opposite of present day photosynthesis which puts reduced material in the ground and maintains an oxidising atmosphere. We might call it 'inverse segregation'.

From geological evidence—for Fe²⁺ in ancient seas, and from the presence of easily oxidised detrital minerals in ancient sediments—the atmosphere is widely thought to have been free of oxygen until ~2,000 Myr ago^{7,8}. That was the great period of ferric sedimentation in the form of the massive and still problematical banded iron formations (BIFs)^{7–10}. Typically these consist of alternate iron-rich and iron-poor layers within siliceous sedimentary rocks. The layers are generally from 0.5–3 cm thick with, frequently, 'microbands' on the mm scale or less that may, nevertheless, extend for hundreds of kilometres: the microbands have been attributed to an annual even outfall from large bodies of water^{11,12}. Could inverse segregation explain these largely oxidised deposits?

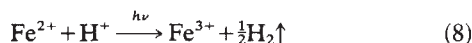
Received 28 August; accepted 31 October 1978.

- Hampson, J. *Can. Air Res. Def. Establ. Tech. Note* 1738 (1966).
- Johnston, H. S. *Science* **173**, 517–522 (1971).
- Crutzen, P. J. *Q. J. R. Meteor. Soc.* **96**, 320–325 (1970).
- Broderick, A. J. *AIAA/SAE 13th Propulsion Conf.* Paper 77-799, Orlando, Florida (1977).
- Oliver, R. C., Bauer, E., Hidalgo, H., Gardner, K. A., & Wasylkiwskyj, W. *US Dept Transportation Rep. FAA-EO-77-3* (1977).
- Duewer, W. H., Wuebblers, D. J., Ellsaesser, H. W. & Chang, J. S. *J. geophys. Res.* **82**, 935–942 (1977).
- Crutzen, P. J. & Howard, C. J. *PAGEOPH* **116**, 497–510 (1978).
- Poppoff, I. G., Whitten, R. C., Turco, R. P. & Capone, L. A. *NASA Ref. Publ.* 1026 (September, 1978).
- Crutzen, P. J. *Geophys. Res. Lett.* **3**, 169–172 (1976).
- McElroy, M. B., Elkins, J. W., Wofsy, S. C. & Yung, Y. L. *Rev. geophys. Space Phys.* **14**, 143–150 (1976).
- Johnston, H. S. *J. geophys. Res.* **82**, 1767–1772 (1977).
- Zahniser, M. S. & Howard, C. J. *4th Biennial Rocky Mountain Regional Meet. Am. Chem. Soc.*, Boulder, Colorado (1978).
- Hudson, R. D. (ed.) *NASA Ref. Publ.* 1010 (1977).
- Turco, R. P. & Whitten, R. C. *NASA Tech. Paper* 1002 (September 1977); *J. atmos. terr. Phys.* **40**, 13–20 (1978).
- Whitten, R. C., Borucki, W. J., Capone, L. A. & Turco, R. P. *Nature* (in the press).
- Howard, C. J. & Evenson, K. M. *Geophys. Res. Lett.* **4**, 437–440 (1977).
- Burrows, J. P., Cliff, D. K., Harris, G. W. & Thrush, B. A. *Phil. Trans. R. Soc.* (in the press).
- Luther, F. M. & Duewer, W. H. *J. geophys. Res.* **83**, 2395–2402 (1978).
- Wofsy, S. C. & McElroy, M. B. *J. geophys. Res.* **78**, 2619–2424 (1973).
- Halocarbons: Effects on Stratospheric Ozone* (National Research Council Report, National Academy of Sciences, Washington, D.C., 1976).

It might account qualitatively for the general history of BIF sedimentation. The record starts with the oldest known sediments at 3,800 Myr ago^{13,14}. As well as carbonates and BIFs, these rocks contain only very small amounts of graphite¹⁴ suggesting that a strongly reducing methane-laden atmosphere, if it ever existed, had passed away by this time¹⁵. As the evolving atmosphere became still less reducing, through a continuing loss of hydrogen to space, together with diminishing volcanic sources¹, reduced species derived from the atmosphere would less often reconvert oxidised precipitates before they reached the sediments. Hence, when the atmosphere was just neutral, inverse segregation should have been at its most effective, corresponding, perhaps, to the greatest period of the BIFs just before they became rare at ~2,000 Myr ago⁸. In the subsequent faintly oxidising regime (until ~700 Myr ago⁷) BIFs could still form, but only in special situations where water masses were fed by submarine sources of Fe²⁺ (refs 8, 10). After that, with substantial photosynthetic oxygen, an ozone screen would have prevented further large scale photoprecipitation. (The BIF record finally peters out at ~800–600 Myr ago¹⁶.)

Quantitatively the idea also seems feasible. If we take the present solar spectral irradiance curve between 200 and 300 nm (ref. 17) and if we assume also a transparent atmosphere in this range, then about 0.03 einsteins yr⁻¹ could have struck an average cm² of sea surface. If each photon lead to the deposition of one Fe³⁺ ion, the sedimentation rate would be ~1,700 mg cm⁻² yr⁻¹. This can be compared with the range of between 8 and 43 mg cm⁻² for total iron (some of it Fe²⁺) found in microbands attributed to an annual outfall¹². Some set of photoredox processes with an overall quantum yield of ~0.005–0.02 would seem to be needed.

Quantum yields of this order are common enough for laboratory photoredox processes. One reaction that is attractive for its simplicity is the well known evolution of H₂ from acid ferrous solutions^{4,18,19}; formally:



Using a low pressure mercury lamp with a filter to cut out radiation below 200 nm, initial quantum yields of about 0.12–0.07 were obtained over the pH range 1–3 with 0.02M FeSO₄ (ref. 4). This may seem too acid for natural waters. It might be objected also that quantum yields fall off as Fe³⁺ builds up⁴. However, the hydrolysis of Fe³⁺:



makes more protons than reaction (8) uses, suggesting possible autocatalysis in surface waters becoming more acid than the sea generally. Reaction (9) should also remove Fe³⁺ from the photoreaction zone by precipitation. (A fresh 0.01M solution of FeCl₃ has a pH of ~2.3 which after 2 days falls to about 1.8 by which time visible precipitation has set in²⁰.)

Getoff's experiments too are suggestive: organic molecules are formed by UV irradiation of CO₂-saturated solutions in the pH range 3.5–8.7 (ref. 6).

Even with the minimum model represented by equations (8) and (9) we can see that particular conditions of Precambrian seas, such as general pH, rates of mixing, and presence of buffers, could have been critical. And there are many other unknowns that could have greatly complicated the situation: traces of ions photoactive at longer wavelengths; a steady-state mix of organic molecules as possible scavengers^{2,5} or photosensitisers¹⁸; or colloidal mineral crystallites allowing surface or solid-state photoreactions²¹. One conclusion is independent of very detailed knowledge: some photoredox processes would surely have been occurring in early Precambrian waters with, sometimes, the formation of volatile reduced species: if the atmosphere remained a sink for these (ultimately through loss of hydrogen to space) then so long as Fe²⁺ was there to prevent the escape from the seas of the corresponding oxidised species, inverse segregation would be a predictable outcome. Whether the banded iron formations are to be explained mainly or

partially in such terms is another matter—but this non-biological light-driven mechanism seems qualitatively and quantitatively feasible.

I thank P. S. Braterman, R. J. Cross and R. Paterson for helpful discussions.

A. G. CAIRNS-SMITH

Department of Chemistry,
University of Glasgow,
Glasgow, UK

Received 18 September; accepted 6 November 1978.

1. Walker, J. C. G., in *The Early History of the Earth* (ed. Windley, B. F.) 537 (Wiley, London, 1976).
2. Fox, M. in *Concepts of Inorganic Photochemistry*, (eds Adamson, A. W. & Fleischauer, P. D.) 333 (Wiley, New York, 1975).
3. Fox, M. F. *Q. Rev.* **24**, 565 (1970).
4. Jortner, J. & Stein, G. *J. phys. Chem.* **66**, 1258, 1264 (1962).
5. Jortner, J., Ottolenghi, M., Rabani, J. & Stein, G. *J. chem. Phys.* **37**, 2488 (1962).
6. Getoff, N. *Z. Naturforsch.* **17B**, 87 (1962).
7. Cloud, P. E. *Science* **160**, 729 (1968).
8. Cloud, P. E. *Econ. Geol.* **68**, 1135 (1973).
9. James, H. L. & Sims, P. K. *Econ. Geol.* **68**, 913 (1973); also 915–1179 (1973).
10. Towe, K. M. *Nature* **274**, 657 (1978).
11. Trendall, A. F. *Econ. Geol.* **68**, 1089 (1973).
12. Trendall, A. F. & Blockley, J. G. *Geol. Surv. Western Australia, Bull.* 119 (1970).
13. Moorbath, S., O'Nions, R. K. & Pankhurst, R. J. *Nature* **245**, 138 (1973).
14. Allaart, J. H. in *The Early History of the Earth*, (ed. Windley, B. F.) 177 (Wiley, London, 1976).
15. Abelson, P. H., *Proc. natn. Acad. Sci. U.S.A.* **55**, 1365 (1966).
16. Goldich, S. S. *Econ. Geol.* **68**, 1126 (1973).
17. Weast, R. C. (ed.) *CRC Handbook of Chemistry and Physics*, 58th edn, F-200 (1977–78).
18. Weiss, J. *Nature* **136**, 794 (1935).
19. Hayon, E. & Weiss, J. *J. chem. Soc.* 3866 (1960).
20. Fryer, J. R., Gildawie, A. M. & Paterson, R. *Nature* **252**, 574 (1974).
21. Fleischauer, P. D. in *Concepts of Inorganic Photochemistry*, (eds Adamson, A. W. & Fleischauer, P. D.) 381 (Wiley, New York, 1975).

Stable isotopic compositions of early Archaean sulphate deposits of probable evaporitic and volcanogenic origins

SIGNIFICANT Archaean bedded sulphate deposits have been described only from the Pilbara Block (Western Australia)^{1,2}, the Barberton Mountainland (South Africa)^{3–6} and southern India⁷. These are all barite deposits, but there is petrographic evidence suggesting replacement of original evaporitic gypsum in the North Pole (Pilbara) and Barberton deposits. This contrasts with the apparent volcanogenic deposition of barite in the relatively sulphide-rich Big Stubby prospect (Pilbara). This letter summarises the petrographic evidence, concentrating on recent data from North Pole, and reports stable isotopic results for barite and associated sulphides which provide corroborative evidence for the environments of formation of these deposits. Some implications concerning the early hydrosphere, atmosphere and biosphere are outlined.

The North Pole deposits^{8,9} (21°07' S, 119°28' E) lie within a ~30-m thick fossiliferous¹⁰ sequence consisting of laminated chert, silicified arenite and conglomerate, between slightly metamorphosed mafic and ultramafic volcanics of the Warawoona Group¹¹. Stratigraphic correlations suggest the North Pole sequence is of similar age to the ~3.45-Gyr-old¹² felsic volcanics of the Duffer Formation, and this is supported by a ~3.4-Gyr 'model age' for North Pole galena (J. Richards, personal communication). Various structures and textures suggest that the bedded barite and chert replaced an original evaporitic gypsum–carbonate sequence^{8,9}. The barite beds are composed of bottom-nucleated crystal groups, 5–20 cm in length, radiating across bedding (Fig. 1). These strongly resemble 'cavoli' and grass-like evaporitic gypsum structures¹³, and they contrast markedly with barite forms from both volcanic-exhalative and primary sedimentary deposits^{14,15}. Some North Pole barite crystals have swallow-tail twins, and where it has been possible to measure interfacial angles of undeformed single



Fig. 1 Section of stratiform barite-chert horizon from North Pole showing large barite pseudomorphs after gypsum with overlying laminated chert (possibly stromatolitic). Universal stage measurements confirm that undeformed barite crystals have gypsum interfacial angles; mean angles for five measured crystals are $110 \wedge 110 = 68^\circ 12' \pm 0^\circ 50'$ and $110 \wedge 111 = 59^\circ 00' \pm 1^\circ 00'$. Most bedded barite crystals contain irregular domains in which barite is slightly misorientated with respect to adjacent domains and to the length of the crystal.

barite crystals, these have indicated crystal faces typical of gypsum but not recorded for barite (Fig. 1). Thus at least some barite is demonstrably pseudomorphing gypsum. Synsedimentary formation of the precursor gypsum is indicated by eroded crystals tops, and sediment drapes and fills between crystals. Intraformational conglomerates containing clasts of bedded barite attest to early lithification of precursor gypsum, if not early baritisation.

Clear, cryptocrystalline silica pseudomorphs after single or twinned gypsum crystals, commonly at high angles to bedding, occur in laminated, fine-grained, grey chert (Fig. 2). Ghost laminae within some of these silica pseudomorphs suggest a diagenetic origin for the crystals. Silica pseudomorphs after cubic minerals, and cubic voids, have also been observed in some chert beds, and a silicified hopper-shaped crystal occurs in one thin section. However, as is also the case in some younger evaporites, there is no evidence for the former presence of abundant halite at the North Pole.

That chert replaced carbonate sediments at North Pole is indicated by relic carbonate which is partially replaced by silica, stylolitic dolomite rhombs, and rhombic voids. Peritidal deposition of chert precursors is implied by laminated (possibly stromatolitic) bedding, graded- and cross-bedding, scour-and-fill structures, intraclast and edgewise conglomerates, ooid grainstones, ripple marks and probable desiccation features such as curved intraclasts and possible mud cracks. Thus depositional environments deduced independently from inter-layered barite and chert are entirely consistent.

Minor bedded pyrite and sphalerite occur near the extremities of the North Pole barite and there are small galena pods within zones of recrystallised barite. Fluid inclusions in barite contain liquid CO_2 , water and traces of H_2S .

The North Pole deposits contrast with the similarly ancient Big Stubby baritic-sulphide deposit in the Duffer Formation, some 40 km to the east. Published descriptions of Big Stubby^{16,17} emphasise its geological, mineralogical and textural similarities to Kuroko-type volcanogenic deposits.

The barite near Barberton^{3,4} ($25^\circ 45' \text{ S}$, $31^\circ 00' \text{ E}$) is in the slightly metamorphosed, fossiliferous¹⁸ Swaziland Supergroup. It is interbedded with carbonated felsic metavolcanics in a cherty sequence at the top of the ~ 3.45 Gyr (ref. 19) Onverwacht Group and in chert and siltstone of the overlying ~ 3.2 Gyr (refs 4, 19) Fig Tree Group. As at North Pole, much of the barite characteristically grew subperpendicular to bedding. It has been noted that it generally resembles evaporites in texture^{4,10}, and some small barite crystals in chert pseudomorph gypsum²⁰. Thus, at least some Barberton barite probably formed in the same way as North Pole barite.

The isotopic results are shown in Fig. 3. $\delta^{18}\text{O}$ values for barite from North Pole and Barberton are both within the range $+4.5$ – $+10.1\%$, except for a single Barberton value of $+13.3\%$. These are the first recorded $\delta^{18}\text{O}$ values for Archaean sulphates and they are, on average, less positive than for Recent evaporitic gypsum ($\sim +9 \pm 13\%$)²¹, possibly supporting the proposal that the Archaean hydrosphere was ^{16}O -enriched relative to younger waters (see ref. 22).

The $\delta^{34}\text{S}$ values for barite from widespread locations at the North Pole (average $+3.6\%$, s. d. 0.5%) and the Barberton deposits (average $+3.8\%$, s. d. 0.4%) are remarkably consistent. They are similar to the results obtained by Perry *et al.*^{2,5} on small sample populations from each area and less variable than the results of Vinogradov *et al.*⁶ for all barite morphologies in the Fig Tree Series (average $+5.0\%$, s. d. 1.5%). Individual post-Archaean evaporitic sulphate deposits also have uniform sulphur isotope composition, but their $\delta^{34}\text{S}$ values are characteristically $\geq +10\%$ (ref. 23); apparently they commonly reflect the isotopic composition of contemporaneous seawater SO_4^{2-} , with $\delta^{34}\text{S}$ values of sulphate minerals being a maximum of 2% heavier than SO_4^{2-} in the parent brine²³. The purportedly volcanogenic barite at Big Stubby has markedly higher $\delta^{34}\text{S}$ values than barite from North Pole and Barberton, and is evidently not reflecting the $\delta^{34}\text{S}_{\text{SO}_4^{2-}}$ values in the contemporaneous hydrosphere.

The $\delta^{34}\text{S}$ values for sulphides from North Pole and from sedimentary rocks penecontemporaneous with the Barberton

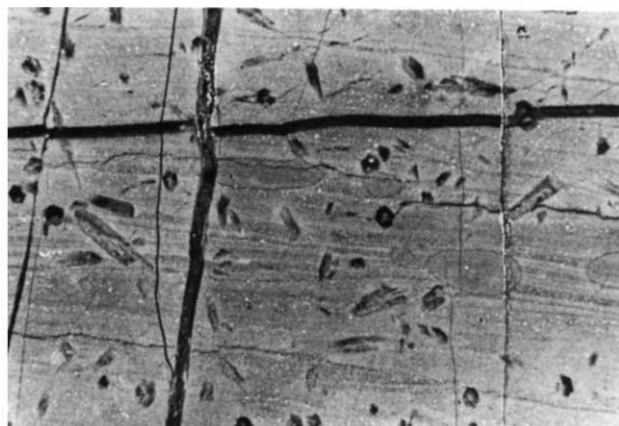


Fig. 2 Photomicrograph of laminated chert (light grey) from North Pole containing well formed silica pseudomorphs after gypsum (medium grey). Pseudomorphs show characteristic interfacial angles for gypsum. Mean angles for 12 crystals are $110 \wedge 110 = 68^\circ 48' \pm 1^\circ 18'$, $110 \wedge 010 = 56^\circ 00' \pm 0^\circ 33'$ and $110 \wedge 111 = 59^\circ 00' \pm 0^\circ 55'$. Field of view 3.5×2.4 cm.

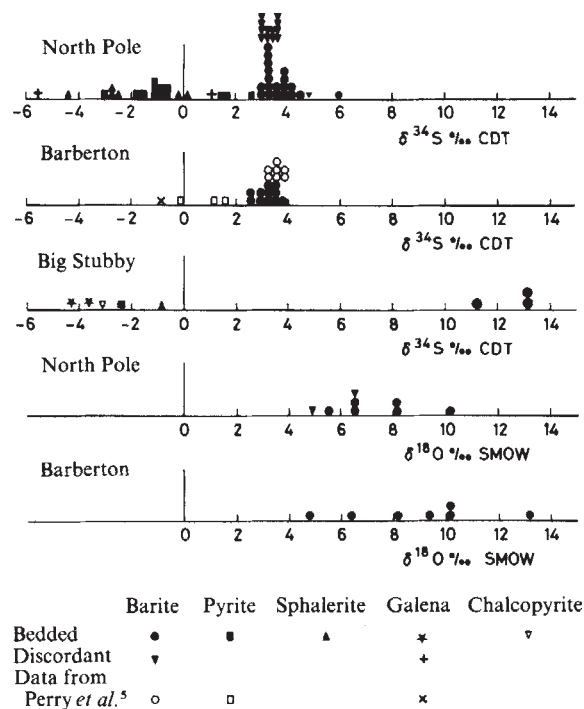


Fig. 3 Stable isotope data. The results from North Pole and Barberton are similar, and they corroborate petrographic evidence for sedimentary–evaporative deposition. Discordant barite at North Pole formed by remobilisation of bedded barite as a result of tectonic doming^{8,9}. The contrasting $\delta^{34}\text{S}$ pattern at Big Stubby supports textural, mineralogical and geological evidence¹⁶ for its volcanogenic origin. Weighted mean $\delta^{34}\text{S}_{\text{SS}}$ values indicate sulphur of juvenile derivation in each case.

barite deposits average -0.9% (s. d. 1.5%), and Big Stubby sulphides have a slightly more negative average $\delta^{34}\text{S}$ value. These are similar to values in sulphate-free, Archaean volcanogenic sulphide deposits²⁴, but differ from younger volcanogenic sulphide deposits which generally do not have $\delta^{34}\text{S}$ values close to 0% (ref. 25).

The weighted mean $\delta^{34}\text{S}$ values for the essentially monomineralic barite deposits at North Pole and Barberton, and for the sulphide-rich Big Stubby deposit, are similar. In each case the $\delta^{34}\text{S}_{\text{SS}}$ in the depositional waters must have been close to 0% , implying sulphur of juvenile origin. The $\Delta^{34}\text{S}$ (barite–sulphide) values of $\sim +15\%$ at Big Stubby are comparable with those in the younger volcano–exhalative deposits²⁵. In contrast, the much smaller $\Delta^{34}\text{S}$ (barite–sulphide) values at North Pole and Barberton (average $< +5\%$) show no evidence of approach to isotopic equilibrium at hydrothermal or surface temperatures²⁶. Ascending juvenile fluids could not have contained $\text{SO}_4^{2-} \gg \text{H}_2\text{S}$, HS^- because sub-surface reactions would have favoured the reduced species. Therefore, the isotopic results indicate that sulphates at North Pole and Barberton could not have been deposited directly from exhalations. These sulphates must have formed largely as a consequence of surficial oxidation of reduced sulphur which was introduced into the hydrosphere by exhalative activity. Oxidation could have been effected by blue-green algae or sulphur bacteria (microfossils are present^{10,18}), although the importance of oxygen production by photolysis of water vapour²⁷ cannot be ruled out. Oxidation of reduced sulphur is a rapid non-equilibrium process which does not lead to significant isotopic fractionation.

The clear evidence for precursor gypsum at North Pole and Barberton argues against deposition of major proportions of primary barite, as proposed by Perry *et al.*⁵. The only model satisfying both isotopic and petrographic constraints is one in which the high concentrations of isotopically homogeneous SO_4^{2-} required for sedimentary gypsum deposition built up in

restricted basins as a result of evaporation and recharge from external sources.

Baritisation of the gypsiferous sediments was presumably accompanied by precipitation of additional barite as the Ba^{2+} -bearing fluids reacted with residual SO_4^{2-} in basin and interstitial waters. Barite produced by these processes would have $\delta^{34}\text{S}$ values similar to precursor gypsum. The origin of the Ba^{2+} is uncertain, but it may have been derived from exhalations related to felsic volcanism (Duffer Formation), or from diagenetic–burial metamorphic reactions.

The minor amounts of sulphides with the barite probably formed from exhalative sulphur which was not oxidised in the depositional basins: the presence of H_2S in fluid inclusions may be significant in this respect. Had they formed by bacterial reduction in SO_4^{2-} -rich evaporative environments, the sulphides would have had wide ranges of $\delta^{34}\text{S}$ values, most being highly negative relative to $\delta^{34}\text{S}$ values of sulphate. Except for the ~ 2.7 -Gyr BP Michipicoten iron formation²⁸, available isotope data for Archaean sulphide mineralisation have not produced evidence for bacterial SO_4^{2-} reduction²⁴.

The occurrence of gypsum and barite does not require an oxygenated atmosphere in the early Archaean; this would only have existed if the oxygen 'sinks' (such as, H_2S , CH_4 , Fe^{2+}) in the hydrosphere and atmosphere were saturated. The sulphate-free volcanogenic sulphides in other, apparently deeper-water Archaean sequences provide evidence that the hydrosphere was not everywhere oxidised. They suggest either the presence of reduced waters under a surficial oxidised zone^{29,30} or the existence of separate seas with different degrees of oxidation.

The Archaean scenario envisaged here is one in which: (1) juvenile, sulphurous exhalations (H_2S , $\text{HS}^- \gg \text{SO}_4^{2-}$) were being added to the hydrosphere and atmosphere; (2) there was local oxidation of reduced sulphur by microbes (or photolysis), but oxygen was not necessarily a stable component of the atmosphere; (3) bedded sulphate deposits formed in shallow-water, evaporitic environments; (4) relatively small, baritic sulphide deposits precipitated directly as a result of volcano–exhalative activity; (5) sulphate-free, exhalative sulphide mineralisation formed in deeper water and/or lower f_{O_2} environments; and (6) bacterial SO_4^{2-} reduction was absent, or not of widespread significance.

We thank R. Binns, J. Glover, M. Muir, D. Sangster, J. Smith, J. Trusswell and M. Walter for critically reading the manuscript, and Dresser Minerals–Getty Oil and Aquitaine Minerals for allowing us to work on their leases. The Baas Becking Laboratory is supported by BMR, CSIRO and the Australian Minerals Industry Research Association.

I. B. LAMBERT
T. H. DONNELLY

CSIRO, Baas Becking Laboratory,
PO Box 378,
Canberra, Australia 2601

J. S. R. DUNLOP
D. I. GROVES

Geology Department,
University of Western Australia,
Perth, Australia 6009

Received 23 May; accepted 27 October 1978.

- Hickman, A. H. *Geol. Survey of Western Australia, A. Rep.* 1972, 57–60 (1973).
- Perry, E. C., Hickman, A. H. & Barnes, I. L. *Geol. Soc. Am. Abstr.* 1226 (1975).
- Viljoen, R. P. & Viljoen, M. J. *Geol. Soc. South Africa, Spec. Pub.* 2, 221–244 (1969).
- Heinrichs, T. K. & Reimer, T. O. *Econ. Geol.* 72, 1426–1441 (1977).
- Perry, E. C., Monster, J. & Reimer, T. *Science* 171, 1015–1016 (1971).
- Vinogradov, Y. I., Reimer, T. O., Leites, A. M. & Smelov, S. M. *Lithol. Miner. Resources* 11, 407–420 (1976).
- Radhakrishna, B. P. & Vasudev, V. N. *J. Geol. Soc. India* 18, 525–541 (1977).
- Dunlop, J. S. R. thesis, Univ. Western Australia (1976).
- Dunlop, J. S. R. *Archaean Cherty Metasediments: Their Sedimentology, Micropalaeontology, Biochemistry and Significance to Mineralization* (eds Glover, J. E. & Groves, D. I.) 30–38 (University of Western Australia, Perth, 1978).
- Dunlop, J. S. R., Muir, M. D., Milne, V. A. & Groves, D. I. *Nature* 274, 676–678 (1978).
- Lippie, S. L. *Geol. Survey of Western Australia, A. Rep.* 1974, 58–63 (1975).
- Pidgeon, R. T. *Earth planet. Sci. Lett.* 37, 421–428 (1978).

13. Richter-Bernburg, G. *Messinian Events in the Mediterranean* (ed. Drooger, W.) 124–149 (North-Holland, Amsterdam, 1973).
14. Igarashi, T., Okabe, K., Yasima, J. *Geology of Kuroko Deposits* (ed. Ishihara, S.) 39–44 (Society of Mining Geologists, Japan, 1974).
15. Shawe, D. R., Poole, F. G. & Brosst, D. A. *Econ. Geol.* **64**, 242–254 (1969).
16. Reynolds, D. G., Brook, W. A., Marshall, A. E. & Allchurch, P. D. *Economic Geology of Australia and Papua New Guinea* (ed. Knight, C. L.) **1**, 185–195 (Aust. Inst. Min. Met. 1975).
17. Sangster, D. F. & Brook, W. A. *Nature* **270**, 423 (1977).
18. Muir, M. D. & Grant, P. R. *The Early History of the Earth* (ed. Windley, B. F.) 595–604 (Wiley, New York, 1976).
19. Jahn, B. M., Shih, C. Y. *Geochim. cosmochim. Acta* **38**, 873–885 (1974).
20. Dunlop, J. S. R. & Groves, D. I. *Archean Cherty Metasediments: Their Sedimentology, Micropalaeontology, Biochemistry and Significance to Mineralization* (eds Glover, J. E. & Groves, D. I.) 39–44 (University of Western Australia, Perth 1978).
21. Fontes, J. C. & Pierce, C. *Abstr. 10th Int. Congr. Sedimentology* 215–216 (1978).
22. Veizer, J. & Hoefs, J. *Geochim. cosmochim. Acta* **40**, 1387–1395 (1976).
23. Holser, W. T. & Kaplan, I. R. *Chem. Geol.* **1**, 93–135 (1966).
24. Donnelly, T. H. *et al. J. Geol. Soc. Aust.* **24**, 409–420 (1978).
25. Sangster, D. F. *Handbook of Strata-Bound and Stratiform ore Deposits* (ed. Wolfe, K. H.) **2**, 219–262 (1976).
26. Ohmoto, H. *Econ. Geol.* **67**, 551–578 (1972).
27. Berkner, L. V. & Marshall, L. C. *J. Atmos. Sci.* **23**, 133–143 (1966).
28. Goodwin, A. M., Monster, J. & Thode, H. G. *Econ. Geol.* **71**, 870–892 (1976).
29. Degens, E. T. & Stoffers, P. *Nature* **263**, 22–27 (1976).
30. Drever, J. I. *Geol. Soc. Am. Bull.* **85**, 1099–1106 (1974).

Metal-enriched sediments from the TAG Hydrothermal Field

A SERIES of sediment samples from the TAG Hydrothermal Field¹ have been collected and analysed for transition metals, and for uranium and thorium isotopes. We show here that metal enrichment of sediment samples is ubiquitous, but geographically variable. Non-detrital metal accumulation rates in one core are Fe and Mn 15.0 and 0.7 mg cm⁻² 10⁻³ yr; and Ni, 180; Co, 93; Cu, 134; and Cr, 148 µg cm⁻² 10⁻³ yr. The hydrothermal field is located on the eastern wall of the Mid-Atlantic Ridge Rift Valley at 26°N and covers ~100 km². It is characterised by extensive deposits of thick, rapidly growing MnO₂ (ref. 2) and by a magnetic low in the axial anomaly¹.

Figure 1 shows that the sediment thickness is quite variable even over a small area of the valley floor. Core 4A (132 cm of sediment) and core 4G (15 cm of sediment) both terminated with basaltic glass, and a core from site 4C recovered glass only, with no overlying sediment. A study of bottom photographs of this section of the median valley³ shows ripple marked sand and other evidence of the strong bottom currents responsible for sweeping sediment off some portions of the valley floor into sediment ponds of uneven distribution.

Three types of sediment were collected during the 1973 cruise of the Discoverer: (1) samples 4A, 4B and 4H are cores of ponded carbonate ooze sediment in the median valley; (2) CaCO₃-rich clay was dredged from the top of a small terrace (dredge site 2A); and (3) sediment interstitial to and covering boulders in the talus was recovered from dredge sites 2A and 3A. This material was less than 5% carbonate and consisted largely of altered basalt debris.

Chemical analyses were made on the sediment samples by atomic absorption spectrophotometry. The results are given in Table 1 along with metal-to-aluminium ratios for average North Atlantic pelagic clay^{4–11} for comparison. The metal-to-aluminium ratios for the average North Atlantic clay is considered to be characteristic of the detrital component of the sediment on which hydrothermal additions are superimposed. Note that none of the samples analysed are very highly enriched in metals compared to the average North Atlantic pelagic clay values. The highest Mn/Al ratios were observed in one of the samples from Dredge Site 2A and the CaCO₃-rich terrace sediment, but the maximum enrichment of Mn relative to the average North Atlantic clay Mn/Al value was only about 7, in contrast to enrichments of 100 or higher found on the East Pacific Rise¹². The presence of a large deposit of Mn oxide found in the nearby hydrothermal field² clearly suggests a significant flux of Mn into seawater in this general vicinity. Scott *et al.*² have measured the accumulation rate of the Mn oxide deposit and found it to be 100–200 mm Myr⁻¹ or 2 orders of magnitude

faster than typical deep sea ferromanganese deposits. It is surprising that so little of the Mn is deposited in the adjacent sediments.

Iron-to-aluminium ratios in the TAG sediments showed even less pronounced enrichment than the Mn/Al ratios. The average Fe/Al ratio for all samples analysed is 0.89 in comparison with the average North Atlantic pelagic clay ratio of 0.56 (Table 1). Most of the Fe in these sediments is detrital in origin. Betzer *et al.*¹³ have found much higher Fe/Al and Mn/Al ratios on a total sample basis for the suspended matter sampled from near bottom water over the ridge crest than for similar particles sampled to either side of the ridge crest. Some of the Fe and Mn in the hydrothermal fluids are probably precipitating in the water column after they enter the ocean. They are diluted in the underlying sediment by biogenic and detrital material being deposited at the same site, and are undoubtedly also being wafted away from the area by currents. The nearly total absence of Fe in the massive hydrothermal MnO₂ deposit² (Fig. 1) suggests that the Fe may be removed as pyrite¹⁴ or other minerals before the hydrothermal fluids debouch onto the sea floor from that vent. The variability in metal enrichments between the massive MnO₂ deposits and the metal enriched sediments indicates variability in the chemistry of the hydrothermal fluids in both space and time.

The metal-to-aluminium ratios of samples from core 4A show that the core is somewhat enriched throughout its length in Fe, Mn, Ni, Cu, Co and Cr with respect to average North Atlantic pelagic clay (Table 1). The Mn content of the 30–32 cm interval falls below the average pelagic clay value, but this is probably the result of dilution by basaltic glass fragments at that depth that were not removed from the sediment before analysis. The Co and Cu values are very strongly correlated with Mn content, showing correlation coefficients of +0.95 and +0.97 respectively. These metals may be concentrated from seawater by precipitating hydrothermal Mn, or they may also be contributed by the hydrothermal solutions themselves. Nickel and Cr in the core are not correlated with Fe, nor with Mn, nor with each other. The factors controlling their distribution cannot be determined without further investigation.

Metal enrichment studies of ridge crest sediments based only on metal-to-aluminium ratios omit a very important parameter—time. The ratios observed in the sediment reflect not only the intensity of the metal source, but also all of the other processes operating to dilute the metals being added by ridge crest processes. An estimate of the rate at which the metals are added compared to analogous rates for non-metalliferous sediments avoids this problem.

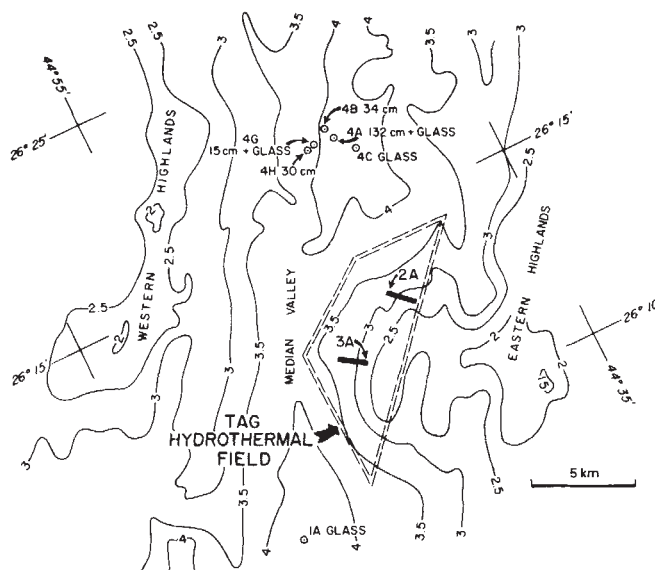


Fig. 1 Sample location map of cores and dredge hauls from the TAG area on the Mid-Atlantic Ridge. Contours are in hundreds of metres.

Table 1 Sediment analyses (CaCO₃-free basis)

	Fe (%)	Mn (%)	Al (%)	Ni (p.p.m.)	Co (p.p.m.)	Cu (p.p.m.)	Cr (p.p.m.)	Fe/Al	Mn/Al	Ni/Al × g ³	Co/Al × 10 ³	Cu/Al × 10 ³	Cr/Al × 10 ³	%CaCO ₃	Porosity (ml cm ⁻³)
2A8-2	6.34	0.145	7.40	128	54.0	47.7	254	0.86	0.0196	1.73	0.730	0.645	3.43	nd*	—
2A8-3	4.26	0.226	2.97	123	59.6	269	242	1.43	0.0761	4.14	2.01	9.06	8.15	nd	—
2A8-4	6.97	1.19	6.63	164	58.2	88.5	342	1.05	0.179	2.47	0.878	1.33	5.16	nd	—
Dredge Site 3A†															
3A2-1	7.47	0.146	8.04	135	61.3	91.3	370	0.93	0.0181	1.68	0.762	1.14	4.60	nd	—
3A2-2	3.15	0.096	3.45	76	42.6	61.2	94	0.91	0.0278	2.20	1.23	1.77	2.72	nd	—
3A2-3	3.73	0.105	3.79	92	35.7	74.0	236	0.98	0.0277	2.43	0.942	1.95	6.23	nd	—
Terrace Sediment‡															
2A8-5	2.90	0.932	3.15	341	140	150	126	0.92	0.296	10.8	4.44	4.76	4.00	77.0	—
Core Tops															
4B02	5.98	0.349	6.74	147	105	166	230	0.89	0.0539	2.18	1.62	2.46	3.41	60.3	—
4G04	5.50	0.528	6.10	204	150	197	208	0.90	0.0866	3.34	2.46	3.23	3.41	72.8	—
4H08	4.89	0.343	5.58	178	108	179	212	0.88	0.0615	3.19	1.94	3.21	3.80	61.2	—
Core 4A															
0-4 cm	5.32	0.407	6.19	218	138	212	244	0.86	0.0658	3.52	2.22	3.41	3.94	67.3	0.73
4-8	5.15	0.421	4.98	198	140	221	204	1.03	0.0845	3.98	2.81	4.44	4.10	69.6	0.74
8-14	3.90	0.478	4.99	229	161	247	180	0.78	0.0958	4.59	3.23	4.95	3.61	73.9	0.75
14-18	3.49	0.419	5.64	212	154	225	169	0.62	0.0743	3.76	2.73	3.99	3.00	72.8	0.71
18-22	4.08	0.472	5.98	235	164	240	184	0.68	0.0784	4.23	2.74	4.01	3.08	75.0	0.73
22-26	4.93	0.441	6.49	206	156	230	190	0.76	0.0680	3.17	2.40	3.54	2.92	73.1	0.75
26-30	5.33	0.254	6.94	261	87.6	146	232	0.77	0.0366	3.76	1.26	2.10	3.34	46.5	0.75
30-31§	4.68	0.205	7.04	276	88.9	442	220	0.65	0.0291	3.92	1.26	6.28	3.12	44.6	—
31-32§	5.53	0.184	8.39	254	81.6	119	256	0.66	0.0219	3.03	0.973	1.42	3.05	31.8	—
32-33§	5.35	0.257	5.01	286	87.8	148	228	1.07	0.0513	5.71	1.75	2.95	4.55	48.6	—
33-37	5.84	0.333	6.12	312	99.9	190	125	0.95	0.0544	5.10	1.63	3.10	2.04	55.9	0.79
120-127	5.86	0.296	6.53	134	79.9	146	212	0.90	0.0453	2.05	1.22	2.24	3.25	51.6	0.77
Avg. N Atlantic pelagic clay ¹⁻¹¹	5.0	0.4	9.0	79	39	115	80	0.56	0.04	0.9	0.4	1.3	0.9	—	—
Mid-Atlantic Ridge 45°N ⁷	9.6	0.5	5.3	214	—	399	128	1.8	0.1	4.0	—	7.5	2.4	—	—
East Pacific Rise ¹²	18.0	6.0	0.5	430	105	730	55	36	12	86	21	146	11	—	—
Avg. Pacific pelagic clay ⁴	5.1	0.9	8.3	300	110	400	78	0.61	0.11	3.6	1.3	4.8	0.9	—	—

* Negligible, not determined. † Material interstitial to talus blocks. ‡ CaCO₃-rich sediment from step-fault terrace. § Most glass shards removed as >0.088 mm fraction. Porosity measurements were made by John Morse.

Table 2 Radiometric data from TAG sediment core 4A

Sample depth (m)	U (p.p.m.)	²³⁴ U/ ²³⁸ U	Th (p.p.m.)	²³⁰ Th <i>d</i> (pm g ⁻¹)	²³⁰ Th/ ²³² Th	²³⁰ Th × <i>s</i> <i>d</i> (pm g ⁻¹)
0-4	0.58 ± 0.03	1.12 ± 0.07	1.38 ± 0.06	14.06 ± 0.40	10.22 ± 0.38	13.58 ± 0.40
18-22	0.75 ± 0.03	1.40 ± 0.07	4.07 ± 0.11	11.60 ± 0.18	11.58 ± 0.27	10.82 ± 0.18
22-26	0.83 ± 0.03	1.37 ± 0.05	6.47 ± 0.44	11.83 ± 0.50	7.43 ± 0.45	10.99 ± 0.50
26-30	0.59 ± 0.03	1.20 ± 0.07	4.55 ± 0.16	10.88 ± 0.24	9.72 ± 0.30	10.37 ± 0.24
33-37	0.73 ± 0.01	1.00 ± 0.02	3.55 ± 0.23	11.32 ± 0.37	12.93 ± 0.77	10.79 ± 0.37
41-50	0.64 ± 0.03	1.26 ± 0.08	6.82 ± 0.19	11.72 ± 0.22	6.98 ± 0.16	11.11 ± 0.22
57-70	0.36 ± 0.04	0.75 ± 0.11	2.51 ± 0.08	8.09 ± 0.11	13.12 ± 0.41	7.89 ± 0.11
120-127	0.65 ± 0.01	1.17 ± 0.03	4.04 ± 0.24	8.27 ± 0.27	8.31 ± 0.46	7.71 ± 0.27

1 s.d. errors from counting statistics are listed.

Table 3 Metal accumulation rates for core 4A

	Fe (mg cm ⁻² kyr ⁻¹)	Mn (mg cm ⁻² kyr ⁻¹)	Al	Ni	Co (μg cm ⁻² kyr ⁻¹)	Cu (μg cm ⁻² kyr ⁻¹)	Cr	CaCO ₃ -free sedimentation rate (mm kyr ⁻¹)
Core 4A, whole sediment	49.5	3.47	61.9	235.0	120.0	214.0	204.0	6.8
Core 4A, detrital fraction	34.5	2.78	61.9	54.8	27.0	80.1	55.6	—
Core 4A, nondetrital* fraction	15.0	0.69	0	180	93	134	148	—
Average N. Atlantic pelagic clay	8.7	0.7	15.6	13.8	6.8	20.2	14.0	2.5
East Pacific Rise	a 82.0	28.0	0.61	160.0	—	—	—	0.5
East Pacific Rise	b 58	18	3	153	47	254	21	5.3

* The accumulation rate for the non-detrital fraction of Core 4A was calculated by subtracting a detrital metal component based on the detrital metal-to-aluminium ratio (average North Atlantic pelagic clay) and the accumulation rate of aluminium in Core 4A.

The sedimentation rate of core 4A was measured by means of the ^{230}Th excess method (Table 2). The rate at which this sediment is accumulating is $\sim 1.8 \text{ cm kyr}^{-1}$. This rate is consistent with Mid-Atlantic Ridge sedimentation rates measured by other workers¹⁵⁻¹⁷. Note in Table 2 that $^{234}\text{U}/^{238}\text{U}$ activity ratios > 1 exist over the whole length of the core, in some cases beyond corrections for 3 s.d. of counting statistics. Post depositional hydrothermal additions of uranium to ridge crest sediments have been previously noted¹⁸⁻²¹. The same interpretation seems reasonable for these data. Addition of ^{234}U -rich uranium would distort the accumulation rate towards a lower value.

Metal accumulation rates have been calculated for core 4A on the basis of average metal concentrations and porosity (Table 1) and the measured sediment accumulation rate of 1.8 cm kyr^{-1} and are listed in Table 3. The Al content of marine sediments is commonly used as an index of the detrital component. We calculated the metal accumulation rates for the detrital component of core 4A using the average Al content of 4A (Table 1), the sediment accumulation rate of 4A, and the metal-to-aluminum ratios for average North Atlantic pelagic clay (Table 1). These values were subtracted from the whole sediment values, yielding the non-detrital accumulation rates listed in Table 3. These rates include only material from non-detrital sources.

The non-detrital metal accumulation rates (Table 3) show that Fe, Mn and the trace metals are being added to sediment in the TAG area at rates that are well above the background values for the North Atlantic Ocean. Iron and manganese accumulation rates on the East Pacific Rise²²⁻²⁴ are much greater than at this site on the Mid-Atlantic Ridge. However, the Ni, Co, Cu, and Cr accumulation rates are conspicuously high in comparison with East Pacific Rise data. The differences observed in sediments from the two active ridge areas may be related to water circulation differences in the hydrothermal systems of fast and slow spreading ridges²⁵. Note also that metal enrichment in TAG area sediments may be higher than average for hydrothermal areas on the Mid-Atlantic Ridge. Sediments from the FAMOUS area at 37°N were found to be accumulating metals at much slower rates than those reported here¹⁸.

This research was supported by the NSF, International Decade of Ocean Exploration, Office of Naval Research and National Oceanic and Atmospheric Administration.

MARTHA R. SCOTT

Department of Oceanography,
Texas A & M University

ROBERT B. SCOTT

Department of Geology,
Texas A & M University,
College Station, Texas 77843

JOHN W. MORSE

Rosensteil School of Marine and Atmospheric Sciences,
University of Miami, 4600 Rickenbacker Causeway,
Miami, Florida 33124

PETER R. BETZER

Department of Marine Sciences,
University of South Florida,
St Petersburg, Florida 33701

LOUIS W. BUTLER
PETER A. RONA

National Oceanic and Atmospheric Administration,
15 Rickenbacker Causeway,
Miami, Florida 33149

Received 3 August; accepted 27 October 1978.

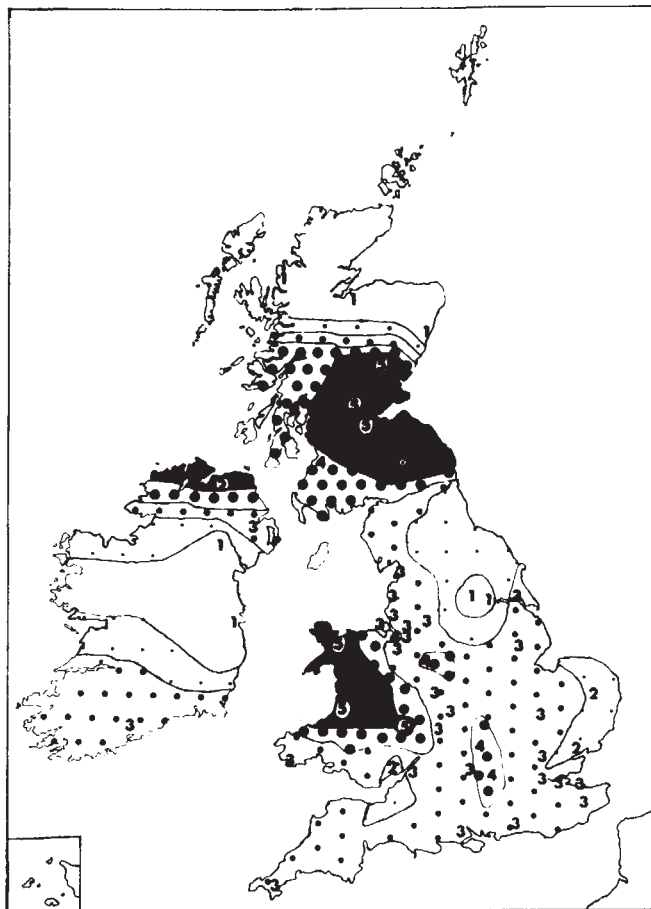
1. Scott, R. B., Rona, P. A., McGregor, B. A. & Scott, M. R. *Nature* **251**, 301-302 (1974).
2. Scott, M. R., Scott, R. B., Rona, P. A., Butler, L. W. & Nalwalk, A. J. *Geophys. Res. Lett.* **1**, 355-358 (1974).
3. Temple, D. G., Scott, R. B. & Rona, P. A. (in the press).
4. Turekian, K. K. & Imbrie, J. *Earth planet. Sci. Lett.* **1**, 166-168 (1966).
5. Cronan, D. S. *Can. J. Earth Sci.* **9**, 319-323 (1972).
6. Bostrom, K. & Valdes, S. *Lithos* **2**, 351-360 (1969).
7. Horowitz, A. & Cronan, D. S. *Mar. Geol.* **20**, 205-228 (1976).
8. Landergrén, S. *Rep. Swed. Deep-Sea Exptn.* **10**, 61-156 (1964).

9. Van der Weijden, C. H., Schuiling, R. D. & Das, H. A. *Mar. Geol.* **9**, 81-99 (1970).
10. El Wakeel, S. K. & Riley, J. P. *Geochim. cosmochim. Acta* **25**, 110-146 (1961).
11. Bruty, D., Chester, R., Royle, L. G. & Elderfield, H. *Nature phys. Sci.* **237**, 86-87 (1972).
12. Bostrom, K. & Peterson, M. N. A. *Mar. Geol.* **7**, 427-47 (1969).
13. Betzer, P. R., Bolger, G. W., McGregor, B. A. & Rona, P. A. *Trans. Am. Geophys. Un.* **55**, 293 (1974).
14. Spooner, E. T. C. & Fyfe, W. S. *Contr. Miner. Petrol.* **42**, 287-304 (1973).
15. Ku, T. L., Bischoff, J. L. & Boersma, A. *Deep-Sea Res.* **19**, 223-247 (1972).
16. Scott, M. R., Salter, P. F. & Barnard, L. A. *Second Maurice Ewing Symp. Vol.* (in the press).
17. Nozaki, Y., Cochran, J. K. & Turekian, K. K. *Earth planet. Sci. Lett.* **34**, 167-174 (1977).
18. Rydell, H. S., Kraemer, T., Bostrom, K. & Joensuu, O. *Mar. Geol.* **17**, 151-164 (1974).
19. Fisher, D. E. & Bostrom, K. *Nature* **224**, 64-65 (1979).
20. Veeh, H. H. & Bostrom, K. *Earth planet. Sci. Lett.* **10**, 372-374 (1971).
21. Scott, M. R. & Salter, P. F. *Trans. Am. Geophys. Un.* **58**, 420 (1977).
22. Dymond, J. & Veeh, H. H. *Earth planet. Sci. Lett.* **28**, 13-22 (1975).
23. Bender, M. et al. *Earth planet. Sci. Lett.* **12**, 425-433 (1971).
24. Bertine, K. K. *Geochim. cosmochim. Acta* **38**, 629-635 (1974).
25. Scott, R. B., Temple, D. G. & Peron, P. R. (in the press).

A survey of ozone levels in the British Isles using indicator plants

THE occurrence of elevated levels of photochemically-produced ozone in the British Isles has been reported by several workers¹⁻⁴. Such elevated concentrations are not restricted to urban areas, but can occur far from sources of pollution^{1,3,5,6}. Most of the measurements of ozone have been made in South-east England; there has been no systematic attempt to study its distribution over the whole country. We report here results of a nationwide survey of ozone concentrations, using the appearance of visible injury on the leaves of a sensitive plant species as an indicator of the presence of the pollutant.

Fig. 1 The distribution of leaf injury, derived by computer interpolation from the mean values of weekly leaf injury index at 53 sites for the period 13 June-5 September 1977. Five classes (denoted by increasing degrees of shading) are distinguished, covering the following ranges of values: 1, 0.01-0.54%; 2, 0.54-1.06%; 3, 1.06-1.59%; 4, 1.59-2.11%; 5, 2.11-2.64%.



The plant used in this study was *Nicotiana tabacum* cv. Bel-W3. This cultivar has a threshold for the appearance of visible leaf injury at 0.04–0.05 p.p.m. ozone (the maximum concentration in which ozone normally occurs naturally), and the extent of leaf injury has been shown to be related to the amount of ozone to which the plant has been exposed^{7–9}. The plants normally were grown in heated glasshouses in standard composts and then 'hardened off' for one week in an unheated glasshouse, before being placed in the field. Their age at this time was 8 weeks. The plants remained in the field for 4 weeks, and were then replaced by a fresh batch of 8-week-old plants. Six replicate plants were exposed at each site. The exposure sites chosen were open, and at a suitable distance from any local source of other pollutants. Wind protection was provided.

The percentage of the area of each leaf covered by typical ozone-induced lesions was estimated visually at each site at weekly intervals. From these data, the change in the mean percentage of leaf area injured (which we term the leaf injury index) was calculated for each week. Photographs and leaf samples were provided to help in identification of ozone damage; while at many sites, plants of an ozone-resistant cultivar, Bel-B, which showed little or no ozone damage, were placed alongside the test plants. Leaves were sent to us to confirm the identification of damage, and to determine the accuracy of damage assessment by comparing the visual estimates with objective estimates obtained using a grid placed over the leaf.

Figure 1 shows the distribution of leaf injury. Because of an absence of data, little importance can be attached to the interpolated values in certain areas (for example, the sharp changes in injury around Berwick). However, it is encouraging that

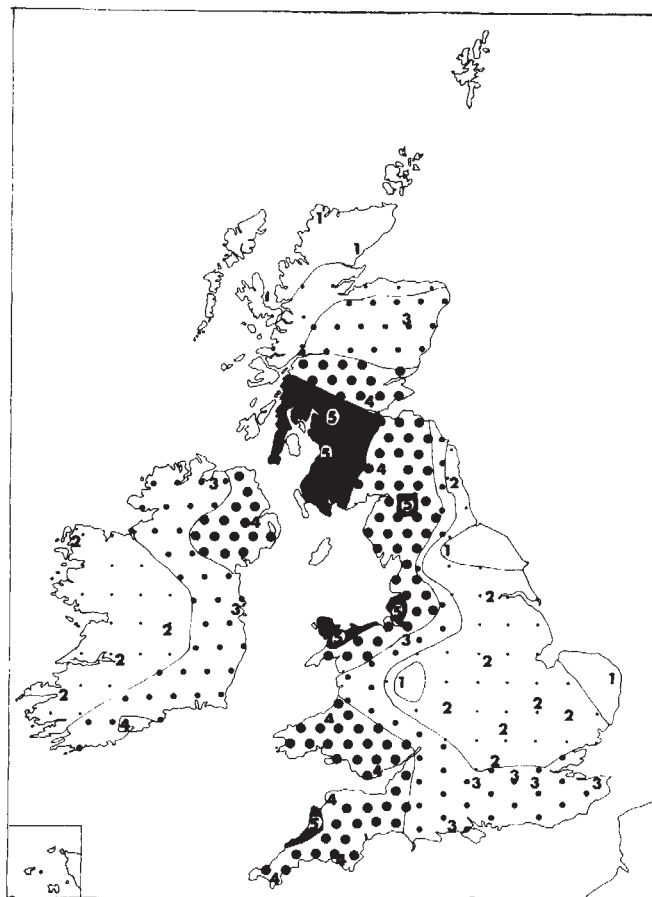


Fig. 2 The distribution of mean values of weekly sunshine hours for the period 13 June–5 September 1977, derived by computer interpolation from Meteorological Office records. The five classes distinguished cover the following ranges of values: 1, 28.9–33.2 h; 2, 33.2–37.5 h; 3, 37.5–41.9 h; 4, 41.9–46.2 h; 5, 46.2–50.5 h.

values for neighbouring sites were usually in agreement. Ozone is produced by a series of photochemical reactions from nitrogen oxides and reactive hydrocarbons. The major source of these pollutants is motor vehicle exhausts, and thus high concentrations of ozone, and large amounts of leaf injury, might be expected to occur in areas of high population density. Yet in 1977, the amount of leaf injury in certain urban areas was relatively low, while it was high in some rural areas, such as North Wales. Only in northern Scotland was no leaf damage found.

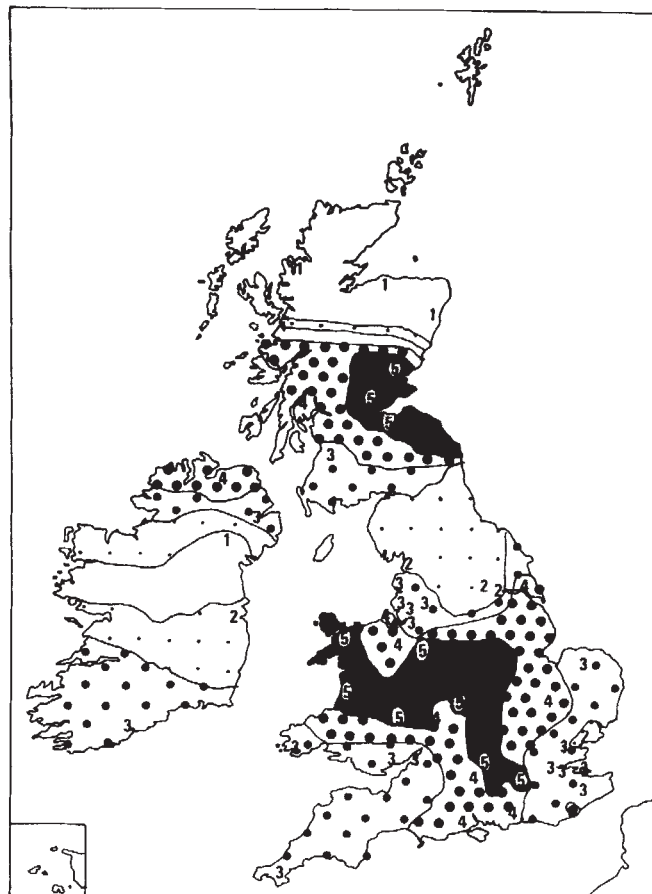


Fig. 3 The distribution of values of k , calculated as described in the text. The five classes cover the following ranges of values: 1, 0.0–0.03; 2, 0.03–0.07; 3, 0.07–0.10; 4, 0.10–0.14; 5, 0.14–0.17.

The levels of sunshine in the summer of 1977 were greatest in the north and west of the country (Fig. 2). As the presence of sunlight is required to drive the reactions leading to ozone formation, its concentrations tend to be high when sunshine levels are high^{2,4,5}. Comparing Figs 1 and 2, it can be seen that the highest amounts of leaf injury tended to occur in those areas receiving the greatest amount of sunshine. We found a significant correlation ($r = 0.554$, $P < 0.001$) between the values of mean weekly leaf injury index (LII) and mean weekly (SH) or daily (S) sunshine hours. This relationship could be expressed by the equation $LII = kSH$ where $SH = \Sigma (S - 6.0)$. The value of 6.0 h gives an estimate of the amount of sunshine needed, on average, to produce sufficient ozone to cause leaf injury. The value of k for the fitted relationship was $0.113\% \text{ h}^{-1}$. Values of k can also be calculated for individual sites. These values can be used to predict the distribution of leaf injury if the amount of sunshine was uniform over the whole country (Fig. 3). This map shows four bands of alternating high and low leaf injury, lying roughly SW–NE across the country, corresponding to regions of high or low degrees of urbanisation. Meteorological Office records show southwesterly and northeasterly to be the predominant wind directions on sunny days in 1977.

The maps shown here are intended to illustrate general patterns—we do not attach great importance to the precise details of Figs 1 and 3 and concede that the model we have used is a very simple one. It is known that meteorological factors other than sunshine affect the magnitude and distribution of ozone concentrations^{1,3-6}. We also concede that the results of one summer must be interpreted with caution; no two summers are the same, especially in Britain. The summer of 1977 was unusually dull, and ozone concentrations were relatively low; stable anticyclonic conditions which are typically associated with very high ozone concentrations^{1,3,5-6} were very rare. However, our results do show that, given favourable weather, ozone concentrations sufficient to damage our test plants can occur anywhere in the British Isles, with the possible exception of northern Scotland. Furthermore it would be expected that both the extent and the concentration of ozone would be lower under the climatic conditions of 1977 than in most years.

We thank all the collaborators who provided these data, and acknowledge the financial support of the NERC.

M. R. ASHMORE
J. N. B. BELL
C. L. REILY

Department of Botany,
Imperial College Field Station,
Silwood Park, Ascot, Berkshire, UK

Received 25 October; accepted 10 November 1978.

1. Apling, A. J. *et al. Nature* **269**, 569–573 (1977).
2. Atkins, D. H. F., Cox, R. A. & Eggleton, A. E. J. *Nature* **235**, 372–376 (1972).
3. Cox, R. A., Eggleton, A. E. J., Derwent, R. G., Lovelock, J. E. & Pack, D. H. *Nature* **255**, 118–121 (1975).
4. Stewart, H. N. M., Sullivan, E. J. & Williams, M. L. *Nature* **263**, 582–584 (1976).
5. Guicherit, R. & van Dop, H. *Atmos. Envir.* **11**, 145–155 (1977).
6. Wolff, G. T., Liroy, P. J., Wight, G. D., Meyers, R. E. & Cederwall, R. T. *Atmos. Envir.* **11**, 797–802 (1977).
7. Bell, J. N. B. & Cox, R. A. *Envir. Pollut.* **8**, 163–170 (1975).
8. Macdowall, F. D. H., Mukammal, E. I. & Cole, A. F. W. *Can. J. Pl. Sci.* **44**, 410–417 (1964).
9. Larsen, R. I. & Heck, W. W. *J. Air Pollut. Control Ass.* **26**, 325–333 (1976).

¹³C Content of human collagen as a measure of prehistoric diet in woodland North America

PLANTS with the C₄ dicarboxylic acid pathway of photosynthesis have markedly higher ¹³C/¹²C ratios than plants with the C₃ or Calvin pathway¹⁻⁴. Cumulative assessments of relative ¹³C concentration ($\delta^{13}\text{C}$) values for C₃ plant foliage yield an average of -26.5‰, relative to the PDB standard whereas C₄ plants average -12.5‰ (ref. 5). These differences in carbon isotope ratios are passed along the food chain to animals and man, so that the nature of the plant food base can be determined through measurements on different body tissues of the consumers. In this study, the time of introduction of maize, a C₄ plant, to the prehistoric inhabitants of several regions in woodland North America, almost exclusively a C₃ plant regime, was determined through isotopic measurements on human bone collagen. The subsequent elevation of maize to the level of a diet staple is also documented.

When plant foods are consumed by animals and man, the ¹³C/¹²C ratios of the plants are further fractionated in the formation of different body tissues^{6,7}. In African browsing ungulates, with diets made up almost exclusively of C₃ foliage, the average $\delta^{13}\text{C}$ measurement (relative to PDB) on bone collagen is -21.2‰ (ref. 8). The collagen is apparently enriched by 5.3‰ relative to the food base, but a possible small C₄ plant component in the diet may have added slightly to this enrichment value. Measurements on other body tissues of the same animals yield average $\delta^{13}\text{C}$ readings of -21.8‰ for hide and hair, -23.5‰ for flesh and -28.9‰ for fat. It can be calculated

Table 1 $\delta^{13}\text{C}$ contents of collagen in adult human skeletons from North American woodland sites, relative to PDB standard

Provenance	Sample size & sex	$\delta^{13}\text{C}$ (‰)	%C ₄
Pre-maize			
1. Koster, Illinois; ~3000 BC	Archaic, 5 M	-21.7 ± 0.34	0
2. Gibson Mounds, Illinois; Middle Woodland, AD ~200	5 M	-20.9 ± 0.63	0
3. Koster Mounds, Illinois; early Late Woodland, AD ~600	5 M	-20.9 ± 1.3	0
4. Du Pont, Ohio; late Archaic, ~2500 BC	5 F	-21.7 ± 0.48	
	5 M	-21.9 ± 1.0	
5. Sand Ridge, Ohio; terminal Archaic, ~1500 BC	1 F	-21.4	0
6. Fairchance, W. Virginia; Middle Woodland, AD ~100–200	5 M	-21.1 ± 1.2	0
Average, pre-maize	31	-21.4 ± 0.78	0
Post-maize			
1. Ledders Mounds, Illinois; late Late Woodland, AD ~1000	5 M	-18.1 ± 2.2	~24
2. Schild Cemetery Knoll B, Illinois; Mississippian, AD ~1200	5 M	-14.1 ± 1.1	~52
3. Noble-Wieting, Illinois; early Upper Mississippian	1	-15.1	~45
4. Turpin, Ohio; Upper Mississippian (Ft. Ancient), AD ~1300	6 F	-12.5 ± 0.55	~64
	4 M	-11.0 ± 1.4	~75
Turpin Site average	(10)	(-11.8 ± 1.3)	(69)

from these values that hunter-gatherers who lived in a C₃ environment such as the North American woodlands and who had diets estimated at 20% animal and 80% plant foods had a food web with an average $\delta^{13}\text{C}$ value closely approximating that of C₃ plants. The observed addition to their diet of freshwater fish and molluscs would not have changed the situation materially as these organisms also resemble C₃ plants in ¹³C content. The addition of C₄ plants to the diet will have a profound effect, of course, and the percentage of C₄ plants in the diet should be measurable in the body tissues. This is also true for marine foods in the diet, as marine fauna have an average $\delta^{13}\text{C}$ value of about -15‰, with values becoming more negative with decreasing ocean temperatures. These predictions can be seen to hold for a sample of 38 pre-Iron Age Europeans, with a food web based on C₃ plants and a possible marine faunal component, and an average $\delta^{13}\text{C}$ value in their bone collagen of -20.4 ± 1.4‰ (refs 9, 10).

To use isotopic measurements as a means of dietary assessment in humans it is necessary to eliminate seasonal factors. Collagen is useful in this respect as its carbon turnover rate is about 10 yr; it is also preserved in archaeological sites. To use collagen, it is also necessary to know accurately the shift in isotopic composition between the plant food base and collagen. We determined this value by measuring the isotopic ratios of collagen from 31 North American skeletons, for an average $\delta^{13}\text{C}$ value (relative to PDB) of -21.4 ± 0.78‰ (Table 1). These skeletons derive from hunter-gatherer cultures of Ohio, Illinois, and West Virginia, with archaeologically documented diets of C₃ plants, terrestrial animals and freshwater fish and molluscs¹¹⁻¹⁶. A few C₄ species of *Amaranthus*, *Panicum* and *Chenopodia* may have been included in this diet, but apparently in insufficient quantity to have a measurable effect. The consistency of the readings and the wealth of archaeological information on diet make this set of readings the most useful to date for calculating the shift in isotopic composition between the base of the food web ($\delta^{13}\text{C}$ = -26.5‰) and human bone collagen ($\delta^{13}\text{C}$ = -21.4‰). The collagen enrichment (+5.1‰) was used in the rest of our study to calculate relative contributions of C₃ and C₄ plants in human diets.

C₄ plants include several cultigens of economic importance, for example, sugar cane, maize and sorghum¹⁷. Their contribution to a diet which is otherwise based on C₃ plants can be

calculated from isotopic measurements on bone collagen. A diet consisting totally of C_4 plants should yield bone collagen with $\delta^{13}C = -12.6 \pm 5.1\%$ or -7.5% . Isotopic measurements between -7.5% and -21.4% reflect a mixture of C_3 and C_4 diet components. Variation between individuals could change the calculated relative percentage by $\pm 6\%$, multiple measurements for a site are, therefore, preferred.

In a pilot study of archaeological skeletons from New York¹⁸, a shift away from the C_3 -based $\delta^{13}C$ average was observed to accompany the introduction of maize cultivation to the area. This effect is confirmed by measurements on 21 skeletons from sites in Ohio and Illinois (Table 1) where the evidence for maize cultivation is clear. It should be noted that the introduction of maize to this region was also accompanied by the exploitation of imported wild C_4 plants (such as amaranth), which cannot be isotopically distinguished from maize. The contribution of these plants is considered to be minor.

The introduction of maize cultivation in the lower Illinois Valley can be assessed from the isotopic readings. There is no measurable C_4 component in the Archaic (Koster), Middle Woodland (Gibson Mounds) or even the beginning of the Late Woodland period (Koster Mound Group). Occasional maize remains occur in the archaeological record from the Middle Woodland onwards, but in quantities small enough to suggest a trade origin. Exotic artefacts from a wide trade network occur in the same deposits. These sporadic injections of maize in the diet were insufficient to register in the long-term isotopic average of bone collagen. Only in the end of the Late Woodland period (Ledders Mounds, ~AD 1000), when fully agricultural Mississippian cultures were already present in other parts of Illinois, can clear evidence for maize cultivation be seen in a C_4 diet component of ~24%. Thereafter, the importance of the cultigen increased rapidly to ~50% by about AD 1200, among populations represented at Schild Cemetery Knoll B (Mississippian) and Noble-Wieting (early Upper Mississippian).

In Ohio, the measurements now available illustrate the wide gap in isotopic values between hunter-gatherers (Du Pont) and fully agricultural, Upper Mississippian peoples (Turpin). The latter show an average C_4 diet component of ~69%. Separated by sex, the averages are ~64% for six females and ~75% for four males. These results suggest a higher C_3 plant intake among women, perhaps in the course of gathering wild plant foods, and are being checked using a larger population. Further measurements are also underway to pinpoint the 'threshold' of maize cultivation in Ohio, Illinois and other Woodland areas.

We thank F. B. Silberbauer and E. Lursen for technical assistance, and the following donors of skeletal material: J. Buikstra (Northwestern University); Cincinnati Museum of Natural History (T. E. Dalbey, R. E. Riggs); R. Schilt (Illinois State University); and West Virginia Geological and Economic Survey (E. T. Hemmings). The research was supported by the University of Cape Town and the Council for Scientific and Industrial Research of South Africa.

NIKOLAAS J. VAN DER MERWE

Department of Archaeology, University of Cape Town,
Rondebosch 7700, South Africa

J. C. VOGEL

National Physical Research Laboratory, PO Box 395,
Pretoria 0001, South Africa

Received 7 August; accepted 6 November 1978.

1. Bender, M. M. *Radiocarbon* **10**, 468-72 (1968).
2. Craig, H. *Geochim. cosmochim. Acta* **3**, 53-92 (1953).
3. Park, R. & Epstein, S. *Geochim. cosmochim. Acta* **21**, 110-26 (1960).
4. Smith, B. N. & Epstein, S. *Pl. Physiol.* **47**, 380-84 (1971).
5. Vogel, J. C., Fuls, A. & Ellis, R. P. S. *Afr. J. Sci.* **74**, 209-215.
6. Stenhouse, M. J. & Baxter, M. S. *Radiocarbon* **18**, 161-72 (1976).
7. De Niro, M. J. & Epstein, S. *Geochim. cosmochim. Acta* **42**, 495-506 (1978).
8. Vogel, J. C. S. *Afr. J. Sci.* **74**, 298-301 (1978).
9. Vogel, J. C. & Waterbolk, H. T. *Radiocarbon* **9**, 107-55 (1967).
10. Vogel, J. C. & Waterbolk, H. T. *Radiocarbon* **14**, 6-10 (1972).
11. Asch, N. B., Ford, R. I. & Asch, D. L. *Illinois St. Mus. Rep. No.* 24 (1972).

12. Buikstra, J. thesis, Univ. Chicago (1972).
13. Featherstone, B. J. *Central St. Anthropol. Soc. Cincinnati* (1972).
14. Hemmings, E. T. W. *Va Archaeologist* **26**, 46-58 (1977).
15. Teeri, J. A. & Stowe, L. G. *Oecologia* **23**, 1-12 (1976).
16. Zawacki, A. A. & Hausfater, G. *Illinois St. Mus. Rep. No.* 17 (1969).
17. Downton, W. J. F. *Photosynthetica* **9**, 96-105 (1975).
18. Vogel, J. C. & van der Merwe, N. J. *Am. Antiq.* **42**, 238-42 (1977).

A C_4 plant from the Pliocene

THE Kranz syndrome^{1,2} is a group of physiological and anatomical features occurring in certain angiosperms; the grasses in particular show increasing efficiency in CO_2 assimilation. Recent research^{3,4} suggests an origin of the syndrome in South America and its subsequent dominance in other continents. The Kranz syndrome occurs in all plants fixing CO_2 by phosphoenolpyruvate carboxylase in the mesophyll and subsequent production of Calvin-Benson cycle intermediates in the bundle sheath⁵. The physiological processes in C_4 plants are closely related to their specialised leaf anatomy known as Kranz anatomy⁶ by which they are distinguished from C_3 plants. It is suggested that ancestral non-Kranz C_3 grasses originated during the Cretaceous period⁴. We report here fossil evidence indicating that C_4 plants occurred at least during the late Tertiary period.

Petrified grasses (Fig. 1a) were collected from two localities (Last Chance Canyon and Sperry Hills area) in the Pliocene Ricardo Formation of California. Several specimens from these localities show similarities in their gross morphological features. The culm, present in the majority of specimens, is notched at the distal end and contains a relatively small amount of mechanical tissue. Conversely, at the base of the round culm there are hypodermal bundles imbedded in abundant sclerenchyma. Sheathing leaf bases are imbricate and wrap around the central culm (Fig. 1d). Wherever a culm is absent the young leaf at the centre is convoluted. Leaves on specimens from Sperry Wash show marked ridges and furrows. These furrows are probably stomatal locations. Thus, these specimens show Kranz anatomy typical of C_4 plants. In addition to a prominent bundle sheath, C_4 species are characterised by a low interveinal distance (less than four mesophyll cells between veins) and the presence of radiating parenchyma in the mesophyll⁷. Mesophyll cells in these fossil leaves are occasionally radiating parenchyma. In the absence of radiating parenchyma up to four mesophyll cells are present between the veins (Fig. 1e). Certain specimens from the Sperry Hills locality have structural features usually associated with C_4 photosynthesis.

Fossil material was subjected to hydrofluoric acid digestion to remove the silicate material. The remainder (mainly cuticle) was then combusted at 800 °C and the CO_2 collected for isotope analysis. The results are presented as $\delta^{13}C\text{‰} = [(R \text{ sample} - R \text{ standard})/R \text{ standard}] \times 1,000$, where R = the mass 45 to mass 44 ratio of sample or standard. The standard was PDB (Chicago). C_4 plants have Kranz anatomy and $\delta^{13}C$ values of -9 to -16‰ while C_3 plants have $\delta^{13}C$ values of -23 to -36‰ and do not exhibit Kranz anatomy. Fossils exhibiting the Kranz anatomy gave a $\delta^{13}C$ value of -13.7‰ , indicative of C_4 photosynthesis, while non-Kranz fossils gave a value of -24.6‰ , similar to C_3 plants.

The existence of the C_4 photosynthetic pathway is supported by the presence in Ricardo Formation of petrified roots resembling extant panicoid plants. These fossil roots have a central stele surrounded by an aerenchymatous cortex (Fig. 1c). Air chambers in the cortex are reticulate as found in *Paspalum commersonii* Lamk. and *P. ciliatofolium* Michx; species of *Paspalum* are known to exhibit C_4 photosynthetic pathways⁸. Previous authors have presented evidence for a fruit resembling *Triticum* in the Eocene Deccan Intertrappean beds of India⁹ and plants resembling the tribes Stipeae and Paniceae¹⁰ in the Miocene-Pliocene beds in North America.

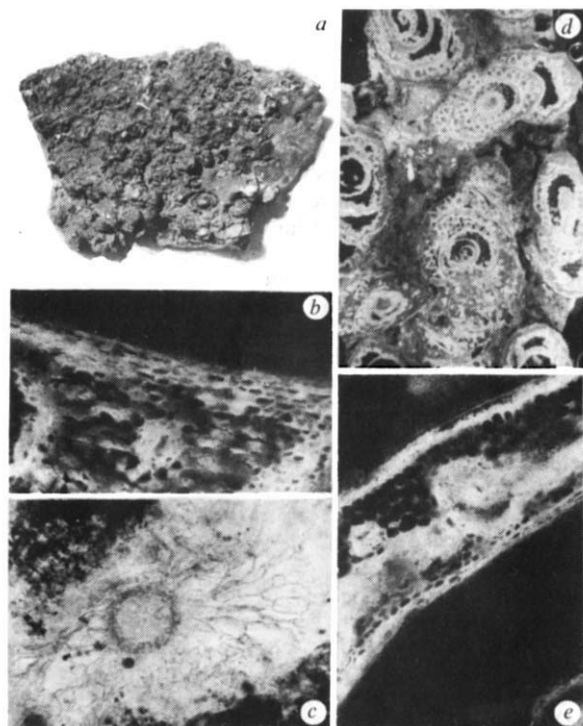


Fig. 1 Petrified grass plants from the Pliocene Ricardo Formation in California. *a*, Distribution of the petrified grass plants in the chert, $\times 1/2$; *b*, radiating parenchymatous cells of the mesophyll tissue occur between veins, $\times 100$; *c*, petrified grass roots with aerenchymatous cortex resembling the extant genus *Paspalum*, $\times 105$; *d*, sheathing leaf bases wrap around a central culm. In the absence of the culm, leaves at the centre are convolute, $\times 17$; *e*, lamina of the leaf in cross section. Between the veins are about 4 mesophyll cells, $\times 29$.

The presence of petrified grass plants in the Ricardo Formation corroborates the reports on grazing animals in these Pliocene beds^{11,12}. Palaeoclimatic determinations are indicative of a plains habitat during Ricardo times¹³ with steep slopes bordering plains and relatively less aridity than at present. Such conditions influenced the growth and propagation of grasses as seen in the Chaparral of Mediterranean climates¹³.

We thank Mr Jim Stichka and K. D. Kehrberg for their help in collecting material.

E. M. V. NAMBU DIRI

Department of Earth Sciences,
University of Manitoba, Winnipeg,
Manitoba, Canada R3T 2N2

W. D. TIDWELL
B. N. SMITH
N. P. HEBBERT

Department of Botany and Range Science,
Brigham Young University,
Provo, Utah 84602

Received 11 September; accepted 8 November 1978.

1. Tregua, E. B., Smith, B. N., Berry, J. A. & Downton, W. J. S. *Can. J. Bot.* **48**, 1209–1214 (1970).
2. Smith, B. N. & Epstein, S. *Pl. Physiol.* **47**, 380–384 (1971).
3. Hartley, W. *Aust. J. Bot.* **6**, 343–357 (1958).
4. Brown, W. V. & Smith, B. N. *Nature* **239**, 345–346 (1972).
5. Hatch, M. D. *CO₂ Metabolism and Plant Productivity*, 59–81 (University Park Press, Baltimore, 1976).
6. Johnson, S. C. & Brown, W. V. *Am. J. Bot.* **60**, 727–735 (1973).
7. Hattersley, P. W. & Watson, L. *Phytomorphology* **25**, 325–333 (1975).
8. Smith, B. N. & Brown, W. V. *Am. J. Bot.* **60**, 505–513 (1973).
9. Chitaley, S. D. & Sheikh, M. T. *J. Indian Bot. Soc.* **50**, 137–142 (1971).
10. Thomasson, J. R. *Science* **199**, 975–977 (1978).
11. Weber, I. E. *Carnegie Inst. Wash. Publ. No. 412*, 115–134 (1933).
12. Merriam, J. C. *Univ. Calif. Publ. Bull. Dept Geol.* **2**, 437a–437c; 438–585 (1919).
13. Axelrod, D. I. *Carnegie Inst. Wash. Publ. No. 476*, 127–164 (1938).

Canine tooth size in female primates

AMONG mammals, weapons such as canine teeth or antlers are often more pronounced in males than in females^{1–5}. Two reasons are normally proposed for such sexual dimorphism. First, there is more competition among males than among females for access to mates and, in consequence, there is enhanced selection for fighting ability^{6–9}. Second, such weapons are important organs of defence for the male, who often assumes the role of defending his mates and offspring from predators^{9,10}. In this paper, we discuss a third factor that is involved: selection acting to influence the size of weapons in females.

In primates, sexual dimorphism in canine size seems to be influenced by both inter-male competition for mates and defence against predators by males⁹. Males of polygynous species, where competition for access to mates is intense, have larger canines (relative to their body size and the canine size of female conspecifics) than males of monogamous species. Also, males of terrestrial species, where anti-predator defence is probably important, have larger canines than those of arboreal species. Much of the variance in canine sexual dimorphism remains unexplained, however, probably partly because additional selective pressures affect female canine size. In particular, selection for enhanced female fighting ability (and thus relatively enlarged female canine size) might be expected in monogamous species, where females participate more frequently in territorial encounters and heterosexual aggression than females of polygynous species^{11,12}. Of the primates that live in larger groups, female aggression may be more frequent in those species that live in multi-male groups than those living in single-male (harem) groups, for two reasons. First, there is growing evidence^{13,14} that female primates may suffer considerable feeding interference from males and it is likely that this will be more intense when there are more males per female in the social unit. (This situation is paralleled in some ungulate species, where females have developed horns or antlers apparently to assist them in feeding competition with males^{15–17}.) Second, although aggressive inter-female competition over resources may occur in either type of group, females in harems are usually more closely related, and so less competitive with each other, than females living in multi-male groups^{7,18,19}.

On these grounds, we predicted that females of monogamous species and those living in multi-male groups should have larger canines relative to their body weight than those living in harem groups. Since there is evidence that terrestrial male primates have relatively larger canines than arboreal species⁹, we

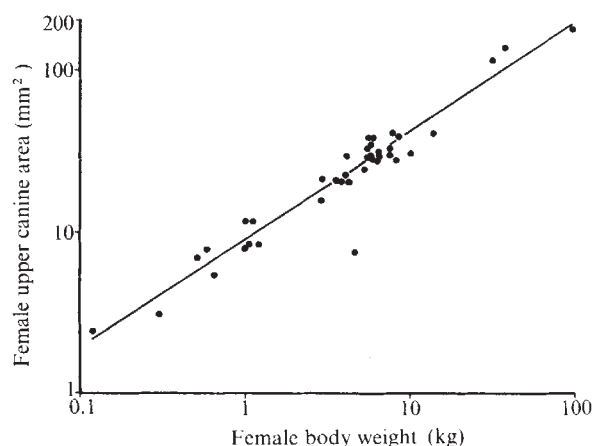


Fig. 1 Female upper canine size regressed on body weight. Both axes are scaled logarithmically.

Table 1 Mean values of relative female canine sizes

	Teeth			
	Upper canine RFCS	<i>n</i> *	Lower canine RFCS	<i>n</i> *
Breeding system				
Harem species	0.818	5	0.861	6
Multi-male species	1.029	15	1.080	14
Monogamous species	1.114	6	1.143	6
Diet				
Frugivores	1.067	25	1.077	24
Folivores	0.926	11	0.945	9
Terrestriality				
Terrestrial species	0.981	8	0.982	8
Arboreal species	1.081	30	1.059	27

* Number of species varies according to availability of data.

examined this relationship in arboreal and terrestrial species separately. Similarly, to allow for any possible effect of diet on female canine size, we have also examined the relationship within folivorous and frugivorous species.

Our measure of canine size is the basal area of the tooth (the product of maximum mesiodistal and maximum buccolingual lengths), which is closely related to canine length ($r = 0.89$, $n = 34$, (ref. 9)). All the dental measurements were extracted from Swindler²⁰ and are presented elsewhere⁹. Data on body weight, diet and breeding systems are taken from Clutton-Brock and Harvey^{21,22}. Female canine size is closely related to body weight (Fig. 1). To remove this effect, we calculated the regression of the natural logarithm of female canine size on the regression of the natural logarithm of non-pregnant body weight for females of 39 species, and measured the deviation in female canine size for each species from the value predicted from female body weight. Our measure of Relative Female Canine Size (RFCS) was the observed canine size, divided by the expected.

As predicted, female primates of harem-forming species have, on average, smaller canines for their body size than do females of other species (Table 1). When the effects of either diet or terrestriality are removed, the differences due to breeding system remain statistically significant for both upper and lower canines (Table 2). There are differences in RFCS between folivorous and frugivorous species, and between arboreal and terrestrial species (Table 1), but when the effect of breeding system is removed, none of these differences is statistically significant (Table 2). In all cases the interaction components are not significant.

Although we have used data from various sources and may therefore have introduced some error²¹, our results give statistical support to the hypothesis that selective pressures for female

Table 2 Statistical analysis of relative female canine sizes according to breeding system, diet and degree of terrestriality

	Teeth			
	Upper canine <i>F</i> _{1,22} statistic	<i>P</i>	Lower canine <i>F</i> _{1,22} statistic	<i>P</i>
Terrestriality and breeding system				
Source of variation				
Terrestriality	1.636	n.s.	0.783	n.s.
Breeding system	20.455	<0.001	4.539	<0.05
Interaction	0.097	n.s.	1.222	n.s.
Diet and breeding system				
Source of variation				
Diet	2.200	n.s.	1.385	n.s.
Breeding system	18.091	<0.001	4.596	<0.05
Interaction	0.136	n.s.	2.923	n.s.

Two-way analysis of variance with unequal sample sizes. Although the sample sizes were the same in each analysis, different species were used, according to the availability of data.

n.s., Not significant.

fighting ability are stronger in monogamous species and those living in multi-male groups than in harem-forming primates.

Two possible artefacts may have affected our results. First, if the group sizes of harem-forming species were significantly smaller than those of other species, then this alone might account for the differences between them because intraspecific competition in the former might be relatively reduced by the lower number of animals present. In fact, harem-forming species do not live in smaller breeding or feeding groups than the other primates in our sample²². The second possibility is that our results are the product of differences between catarrhine and other primates; catarrhines are noted for their well-developed canines^{1,23,24}. It is unlikely, however, that this could explain the relationship between breeding system and RFCS, since all the species in our sample of harem-forming species were catarrhines.

We thank Drs S. Blaffer Hrdy and R. D. Martin and M. Hall for helpful comments.

PAUL H. HARVEY*

MICHAEL KAVANAGH

School of Biological Sciences,
University of Sussex,
Brighton, Sussex, UK

T. H. CLUTTON-BROCK

King's College Research Centre,
Cambridge, UK

Received 10 July; accepted 23 October, 1978.

* Present address: Departments of Biology and Anthropology, William James Hall, Harvard University, Cambridge, Massachusetts 02138.

1. Crook, J. H. in *Sexual Selection and the Descent of Man, 1871-1971* (ed. Campbell, B.) 231-281 (Aldine, Chicago, 1972).
2. Geist, V. *Mountain Sheep: A Study in Behavior and Evolution* (Chicago University Press, Chicago, 1971).
3. Lorenz, K. *On Aggression* (Methuen, London, 1963).
4. Darwin, C. *The Descent of Man and Selection in Relation to Sex* (John Murray, London, 1871).
5. Trivers, R. L. in *Sexual Selection and the Descent of Man, 1871-1971* (ed. Campbell, B.) 156-179 (Aldine, Chicago, 1972).
6. Struhsaker, T. T. *Folia primat.* **11**, 80-118.
7. Clutton-Brock, T. H. & Harvey, P. H. in *Growing Points in Ethology* (eds Bateson, P. P. G. & Hinde, R. A.) (Cambridge University Press, London, 1976).
8. Wilson, E. O. *Sociobiology, the New Synthesis* (Belknap, Harvard, 1975).
9. Harvey, P. H., Kavanagh, M. & Clutton-Brock, T. H. *J. Zool.* (in the press).
10. DeVore, I. & Hall, K. R. L. in *Primate Behavior: Field Studies of Monkeys and Apes* (ed. DeVore, I.) (Holt, Rinehart and Winston, New York, 1965).
11. Kleiman, D. G. *Q. Rev. Biol.* **52**, 36-69 (1977).
12. Clutton-Brock, T. H. & Harvey, P. H. *Nature* **273**, 191-195 (1978).
13. Packer, C. R. & Pusey, A. E. *Folia primat.* (in the press).
14. Dittus, W. P. J. *Behaviour* **63**, 281-322 (1977).
15. Geist, V. *Am. Zool.* **14**, 205-220 (1977).
16. Chapman, D. I. *Mam. Rev.* **5**, 121-172 (1975).
17. Whitehead, G. K. *Deer of the World* (Constable, London, 1972).
18. Blaffer Hrdy, S. *The Langurs of Abu: Female and Male Strategies of Reproduction* (Harvard University Press, Harvard, 1977).
19. Seger, J. in *The Langurs of Abu: Female and Male Strategies of Reproduction* (by Blaffer Hrdy, S.) 317-326 (Harvard University Press, Harvard, 1977).
20. Swindler, D. R. *Dentition of Living Primates* (Academic, London, 1976).
21. Clutton-Brock, T. H. & Harvey, P. H. in *Primate Ecology: Studies of Feeding and Ranging Behaviour in Lemurs, Monkeys and Apes* (ed. Clutton-Brock, T. H.) (Academic, London, 1977).
22. Clutton-Brock, T. H. & Harvey, P. H. *J. Zool.* **183**, 1-39 (1977).
23. Napier, J. R. & Napier, P. H. *A Handbook of Living Primates* (Academic, New York, 1967).
24. Hill, W. C. O. *Evolutionary Biology of the Primates* (Academic, London, 1972).

The distance and nature of the light-trap response of moths

LIGHT TRAPS of various forms have been used to collect and study moths for well over 100 yr, but surprisingly little is known about how they attract moths. There has been some evaluation of the factors influencing the size of light trap catches¹⁻⁵ and of the mechanics of the terminal phase of the moth's approach to a light⁶, but virtually nothing is understood about the light-trap response itself. Such an understanding is perhaps unnecessary when light traps are used solely to collect specimens, but becomes crucial as soon as they are used for quantitative sampling or survey work⁷. Of particular importance to the interpretation of such work is a knowledge of the distance from which

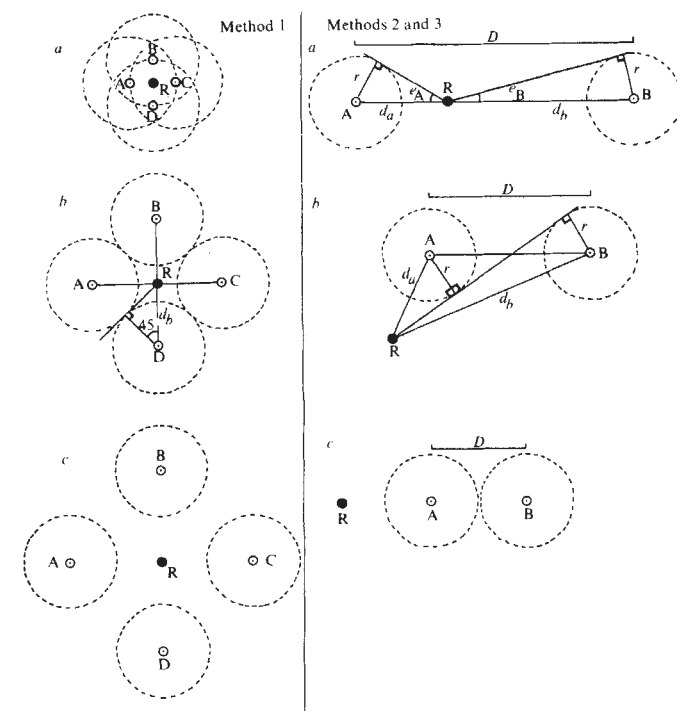


Fig. 1 Three mark-release/recapture methods for estimating the distance of response of moths to a light trap; 1—the increasing circle method. *a*, Four light traps are placed equidistant from a release point (*R*) at such a distance that the release point is within the response range (dashed circles) of moths to the light traps. The number of marked moths captured by the traps should be relatively high. *b*, As the four traps are moved further and further away from *R*, there comes a position first at which the edge of the response range withdraws from the release point and second, at which the edges of the response ranges of the four traps cease to overlap but just touch. Even at this position, however, the released moths cannot escape the circle of traps without entering a response range. The proportion of recaptured moths should therefore be the same as for *a*. *c*, Movement of the traps still further from *R* allows some moths to escape the circle of light traps by flying through the gaps between the response ranges. The proportion of moths recaptured should decrease as the distance of the traps from the release point increases. See Fig. 2(i) for the expected proportion of moths recaptured against distance of the traps from *R*. The radius of the response range is then given by $0.707 d_b$ (where 0.707 is the sin of 45° and d is the distance of each trap from the release point at position *b*). 2—The angle ratio method. *a*, When *R* is nearer one trap (*A*) than another (*B*), the angle (e_A) subtended to *R* by the response range of *A* is greater than the angle (e_B). Assuming that individuals leave *R* equally in all directions, the ratio (a/b) of marked moths recaptured in *A* and *B* should be e_A/e_B . This ratio is a function of r , the radius of the response range ($\sin e_A = r/d_a$; $\sin e_B = r/d_b$). For any given value of e_A/e_B the value of r can be determined by trial and error. 3—The masking method. Starting from the position used for method 2, let the distances d_a and d_b remain constant but let the distance (D) between the two traps decrease (*b*). As D decreases, the release point *R* swings around trap *A*. The ratio of marked moths recaptured in *A* and *B* should remain constant. This remains true even at the position shown (*b*) where a single line drawn from *R* is tangential to the response ranges of both *A* and *B*. Further decrease in D causes the response range of trap *B* to be increasingly masked by the response range of trap *A*. The minimum distance possible for D and the maximum masking of *B* by *A* is reached at *c*. See Fig. 2*b* for the expected shape of the curve of a/b plotted against D . The point of inflection locates *b* from which, given d_a , d_b and D , r can be calculated:

$$r = \frac{4d_a^2D^2 - (d_b^2 - d_a^2 - D^2)^2}{8d_b^2 + 8d_a^2 - 4D^2}^{1/2}$$

Methods 2 and 3 therefore give two independent estimates of r : from the graph, the value (V) of a/b along its horizontal portion gives one estimate (method 2); the point of inflection locates b and gives another estimate (method 3).

moths orientate with respect to a light source; it seems intuitively that this distance should be fairly large. We present here the results of three experiments designed to determine the distance of response of free-flying moths to an artificial light source. Our results support Sotthibandhu's claim⁴ that the effective range of a 125 W mercury vapour (MV) lamp is about 3 m. They also lead to speculation concerning the behavioural meaning of the light trap response in moths.

There was general agreement that the distance of response to a 125 W MV lamp was of the order of 0.5 km (on a moonless night) (see ref. 8 for review). The first attempt to measure the distance at which individual moths in the field orientate towards such a lamp (positioned 60 cm above ground level) was made by Sotthibandhu⁴ who claimed that the distance was of the order of only 3 m. Using flying large yellow underwing moths, *Noctua pronuba*, tethered to apparatus that allowed flight direction to be monitored continuously, he showed that the distance of response to a ground-based light trap varied between 1 and 4 m and was a function of environmental variables such as cloud cover, ambient temperature, and Moon illumination. Further-

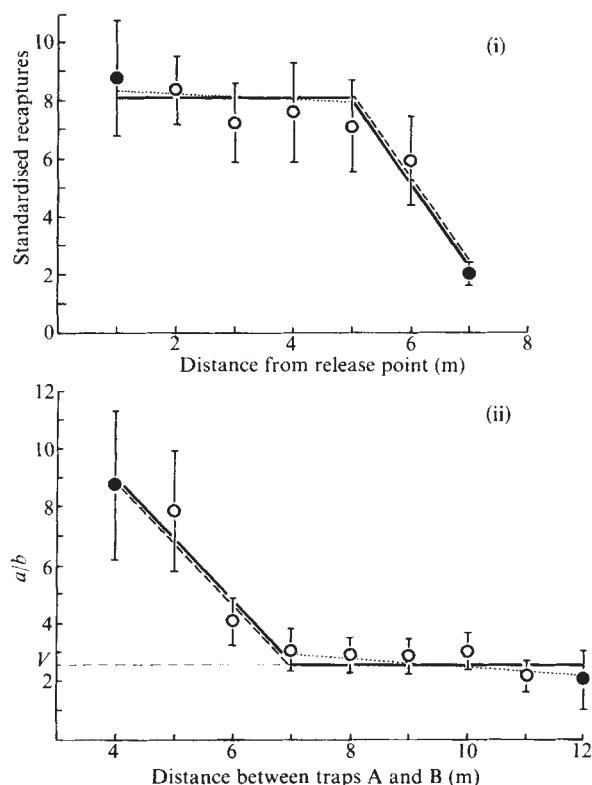


Fig. 2 Analysis of results presented in Tables 1 and 2. Experiments were carried out between 26 June and 26 July 1978 on a lawn at Woodchester Park Field Centre, Glos. Two species of moths, the large yellow underwing, *N. pronuba*, and the heart and dart, *Agrotis exclamationis*, were used. Moths to be marked and released were captured the previous night by the experimental light traps. After marking, without anaesthetisation, with a single dot of cellulose paint on the wing, the moths were stored (in communal keep nets in a cool room) until experimental release at about 22.30 GMT. The moths were released onto the ground within a 10-cm radius of the release point. Dots plus bars indicate means $\pm$ s.e.m. $\bullet$, means; $\circ$, sliding means for the distance indicated ± 1 m. All regression analyses are for the raw data shown in Tables 1 and 2, not for the means and sliding means. (i) and Table 1 suggest (method 1) that the recaptures start to decrease beyond a release distance of 5 m. Calculated regression lines (dotted and dashed lines) are not significantly different ($P > 0.10$) from the expected (solid) lines. It may reasonably be concluded that the data do not depart significantly from the shape (solid line) predicted by the theory on which the method is based (Fig. 1), given that the response range radius is 3.5 m (that is, 0.707×5). (ii) Similarly, the data obtained by the masking experiment (Table 2, Methods 2 and 3) do not depart significantly from the shape predicted (solid line) by the theory on which the method is based (Fig. 1), given that a point of inflection occurs at a value for D of 7 m. The value (V) for a/b over the horizontal part of the curve is 2.09. This is obtained from Table 2 to avoid the effects of Bailey's correction for small numbers. V is calculated for $D \geq 7$ by taking the total number of marked moths recaptured in *A* ($a = 147$) and the total number of marked moths recaptured in *B* ($b = 42$), correcting for the total number of unmarked moths captured in each (1096 in *A*, 654 in *B*), giving $a = 0.134$ and $b = 0.064$, and then dividing a by b . Given that in our experiment $d_a = 4$ m and $d_b = 8$ m, this ratio of 2.09 gives a response range radius of about 2.2 m (Method 2). Given $d_a = 4$ m and $d_b = 8$ m, a point of inflection at $D = 7$ gives a response range radius (r) of 2.7 m (Method 3). Summarising: the three estimates of response range radius are: method 1, $r = 3.5$ m; method 2, $r = 2.2$ m; method 3, $r = 2.7$ m. The three independent estimates of response range given by the three methods are therefore consistent both with each other and with the view that the two moth species studied respond only from a short distance of about 3 m.

Table 1 Effect of varying the distance of release from four equally spaced light traps on the recapture of marked moths (method 1)

Capture/recapture data for four traps (A, B, C, D)														
d_b	date	M	A		B		C		D		A	B	C	D
			$r+1$	N	$r+1$	N	$r+1$	N	$r+1$	N	R	R	R	R
1	6/7	165	2	8	3	21	3	25	2	29	15.15	8.66	7.27	4.18
2	5/7	143	5	27	3	20	5	47	4	58	12.95	10.49	7.44	4.82
3	7/7	98	2	19	3	111	1	36	1	8	10.74	2.76	2.83	12.76
4	16/7	175	2	27	6	175	2	89	2	8	4.23	1.96	1.28	14.29
5	8/7	177	8	24	14	183	1	21	1	4	18.83	4.32	2.69	14.12
6	22/7	198	6	56	4	43	5	28	2	24	5.41	4.70	9.02	4.21
7	15/7	293	12	118	2	45	2	38	2	49	3.47	1.52	1.80	1.39

d_b , Distance of light traps from release point in metres; M , number of marked moths released; r , number of marked moths recaptured (corrected by Bailey's correction for small samples¹⁶ to give $r+1$); N , number of unmarked moths captured; R , standardised recaptures (that is, number that would have been recaptured if for every night $M=100$ and $N=100$), given by $R=[(r+1)100 \times 100]/M.N$. Standardisation was necessary because multiple regression analysis showed the number recaptured to be a significant function ($P<0.02$) not only of the number released but also of the number of unmarked moths captured.

Table 2 Effect of varying the distance between two traps on the relative proportions of marked moths recaptured in each (method 3)

D	Date	M	Trap A				Trap B			
			r	N	$a=(r+1)/N$		r	N	$b=(r+1)/N$	a/b^*
4	30/6	227	49	87	0.575		5	55	0.109	5.27
4	1/7	282	12	50	0.260		2	69	0.043	5.98
4	18/7	174	25	53	0.491		1	61	0.033	14.96
6	9/7	251	38	375	0.104		7	399	0.020	5.19
7	13/7	290	35	457	0.079		3	147	0.027	2.89
8	2/7	415	11	31	0.387		14	118	0.127	3.05
8	3/7	403	38	34	1.147		3	4	1.000	1.15
9	14/7	428	27	282	0.099		5	275	0.022	4.55
10	21/7	65	13	118	0.119		0	22	0.045	2.61
11	24/7	139	6	8	0.875		6	16	0.438	2.00
12	28/6	97	10	156	0.071		3	33	0.121	0.58
12	29/6	279	7	10	0.800		8	39	0.231	3.47

Trap A, 4 m from release point (R); Trap B, 8 m from R; D , distance between trap A and B in metres; M , number of marked moths released; r , number of marked moths recaptured; N , number of unmarked moths captured.

* Calculated from a and b before they are rounded to 3 decimal places.

more, for each individual as it was brought nearer to the light, the change-over from celestial orientation⁹ to light-trap orientation occurred suddenly at some threshold distance. The implications of such a short threshold response distance of moths to a light trap are far-reaching, particularly with respect to the use of light traps in survey work^{2,7}, in that, if it is correct, shifts of only a few metres in the position of a light trap could markedly influence the number and species of moths captured. However, it is inevitable that some initial reservations are felt over Sothibandhu's results because he used tethered moths. There is thus a need for the short-distance nature of the light-trap response to be confirmed using free-flying individuals.

The theory behind our three experimental methods, all of which are based on mark-release/recapture, is presented in Fig. 1. The field procedure is described in the legend to Fig. 2 and the results are presented in Tables 1 and 2. The results are analysed and interpreted in Fig. 2. During the course of this work nearly 5,000 individuals of two species of moth were marked and released. The results (Fig. 2) for all three methods are consistent with the suggestion that moths only respond to a 125 W MV lamp from short distances (about 3 m). Why should moths respond to an artificial light source from such a short distance when it is clear from electroretinographic studies¹⁰ that the light can be perceived from much greater distances?

It has been shown⁹ that the angle of orientation (which is individual-specific and which may be anything within the range $0^\circ \pm 90^\circ$ for individuals caught in a light trap) of an individual *N. pronuba* relative to an artificial light source is the same as the angle taken up by that individual relative to the Moon's azimuth. The implication is, therefore, that the light response of moths is a Moon-orientation response, a suggestion first made by von

Buddenbrock¹¹ but which is no longer favoured generally^{12,13}. If Buddenbrock was right, the problem remains of why the moths should respond to the Moon at a distance effectively of infinity whereas they respond to an artificial light source at a distance of only 3 m.

Using tethered moths, Sothibandhu has shown⁴ that the response distance to a 125 W MV lamp raised 9 m above ground level is much greater than for the same lamp only 0.6 m above ground level, being of the order of 10–17 m. The angle of elevation of the light source is therefore important, as would be expected if the light-response is a Moon-orientation response. When the light source is near the ground, it is below the height of most flying moths and individuals are often seen to fly overhead, a few metres above ground level, without apparently responding to the light. Others are seen suddenly to dive down to ground level. This behaviour is often a response to bats flying overhead¹⁴ but also frequently occurs in the apparent absence of bats. If the diving response is a response to the light, it could be interpreted as an attempt to restore a Moon substitute, the artificial light, to an acceptable elevation, above the level of the moth's eyes, so that dorsal rather than ventral ommatidia are illuminated⁹. Once a moth is near or on the ground, of course, the elevation of the light can approach that of the real Moon, depending on the distance of the moth from the light. Perhaps the explanation for moths responding to a light 60 cm above ground level, at a distance of only 3 m is that at greater distances, even for low-flying moths, the apparent elevation is not a close enough approximation to that of even a newly risen Moon.

Angle of elevation cannot be the only factor involved in Moon recognition, however, otherwise, when the artificial light is raised 9 m above ground level the distance of response would be

expected to be much greater than the 10–17 m observed. Spectral composition, light intensity and the physical size of the light are all factors that could also be important. The light bulbs used by us, which are the same as those used by Sotthibandhu⁴, have a horizontal diameter of 7.5 cm and a vertical diameter of 12 cm. As the Moon changes phase its apparent horizontal diameter changes considerably but the vertical diameter is much more constant, subtending to a point on the Earth's surface an angle of 0.518° (diameter = 3,456 km; distance = 382,170 km). The vertical dimension of the light bulb used by us subtends the same angle to a point at a distance of 13.27 m. This is much nearer the distance of response (10–17 m) of a moth near the ground to an elevated (9 m above ground level) light source than is the distance (35–519 m, depending on Moon phase, latitude, and so on⁸) at which the light source has the same level of illumination as the Moon.

We suggest, therefore, that the light-trap response of moths results when the moth ceases to recognise that the artificial light is not the Moon, and that the most important features used in this recognition are first, the elevation, and second, the vertical diameter of the light source. Level of illumination and spectral composition may also be involved^{5,8} but are suggested here to be far less important.

On the interpretation of the light-trap response offered here, light traps should catch only those individuals that are in the physiological state in which they would make use of Moon orientation. Most often these will be individuals that are in the process of, or just initiating or terminating, migration. This could explain why light traps do not capture all species of moths in the vicinity² (non-migrants not showing the light-trap response) and why males of most species predominate in light-trap catches, males probably being more migratory than females¹⁵. It could also explain why a better correlation is obtained between light trap catch and volume of migration as determined by radar³ than between light-trap catch and suction-trap catch².

R. ROBIN BAKER
YVONNE SADOVY

Department of Zoology,
University of Manchester,
Manchester, UK

Received 25 August; accepted 6 November 1978.

- Williams, C. B. *Phil. Trans. B* **244**, 331–378 (1961).
- Taylor, L. R. & Carter, C. I. *Trans. R. ent. Soc., Lond.* **113**, 369–386 (1961).
- Schaefer, G. W. in *Insect flight, Symp. R. ent. Soc. Lond. no. 7* (ed. Rainey, R. C.) 157–197 (Blackwell, London, 1976).
- Sotthibandhu, S. thesis, Univ. Manchester (1978).
- Mikkola, K. *Ann. Zool. Fennica* **9**, 225–254 (1972).
- Hsiao, H. S. *Attraction of moths to light and to infrared radiation* (San Francisco Press, 1972).
- Taylor, L. R., Kempton, R. A. & Woiod, J. P. *J. Anim. Ecol.* **45**, 255–272 (1976).
- Bowden, J. & Morris, M. S. *Bull. ent. Res.* **65**, 303–348 (1975).
- Sotthibandhu, S. & Baker, R. R. *Anim. Behav.* **27** (in the press).
- Agee, H. R. *Ann. Ent. Soc. Am.* **65**, 701–705 (1972).
- Buddenbrock, W. von *Grundriss der vergleichenden Physiologie* (Berlin, 1937).
- Callahan, P. S. *Nature*, **206**, 1172–1173 (1965).
- Laithwaite, E. R. *Entomologist* **93**, 133–137 (1960).
- Roeder, K. D. in *Animal Behavior: Readings from Scient. Am.* (Eds Eisner, T. & Wilson, E. O.) 46–54 (W. H. Freeman, San Francisco, 1975).
- Baker, R. R. *The Evolutionary Ecology of Animal Migration* (Hodder and Stoughton, London, 1978).
- Bailey, N. T. J. *J. Anim. Ecol.* **21**, 120–127 (1952).

Male mating experience and competitive courtship success in *Drosophila melanogaster*

MULTIPLE copulations by *Drosophila melanogaster* males reduce their fertility¹, even though these males will continue to court and mate after four or five successive matings. This sterility involves depletion of the accessory glands and not of sperm

supply² and is only temporary, for when these males are mated 2 or 3 h after the onset of sterility, they once again transfer sperm. A female *Drosophila* would maximise her reproductive fitness by mating with a male which transfers the most sperm; however, virgin females will copulate with temporarily sterile males when placed in a single-pair situation. Although female *Drosophila* are capable of discriminating mates on the basis of genotype differences³, it is not known whether they will selectively copulate with fertile males in competition with temporarily sterile males of the same strain. In the experiments reported here, we have examined the relative mating success of experienced males with temporarily reduced fertility and virgin males from the Canton-S strain of *D. melanogaster*. The data show that when a virgin female is placed with a virgin and an experienced male, she tends to mate with the virgin male. Examination of some possible factors for increased virgin male success, such as time until initiation of courtship, amount of time spent courting, and overall mating speed, showed that for the courtship parameters measured, virgin and experienced males are equal.

Flies were raised on standard cornmeal-molasses-agar medium at 24 ± 1 °C. Virgin flies were separated by ether anaesthesia and stored until 3 d old. Virgin females were then placed individually into 8-dram shell vials containing one experienced male and one virgin male. Males were distinguished by wing clipping³, which was done at the time of collection to avoid the need for a second anaesthesia. Clipping was alternated between groups to avoid any possible effects on the results. Small clips have not been found to alter mating in the present study (see below) or in previous studies⁴. All subsequent handling of flies was done by aspiration.

Table 1 Number of offspring produced by males mating three successive times

Mating	Mean no. of offspring per mating per male*	No. of males
1	288.0 ± 17.4	33
2	296.1 ± 28.2	31
3a	194.7 ± 23.6	30
3b	206.0 ± 21.0	29
3c	303.2 ± 18.8	35

a, Males mated immediately after the second mating; b, males rested 1 h after the second mating; c, males rested 24 h after the second mating.

* A Duncan multiple range test places these data in two statistically significantly different subsets ($P < 0.01$) where offspring number from mating 1, 2 and 3c are in one subset, and offspring numbers from matings 3a and 3b are in another subset.

Three series of experiments were carried out. In series 1, 'experienced' males were obtained by placing virgin males in a shell vial containing one virgin female. Immediately after copulation, the male was placed with another virgin female. During the initial phase of the experiments, 97% of the original males had mated the first and second times by the 1-h cut-off period. After the second mating, each experienced male was placed with a virgin male in a clean shell vial, and a virgin female was introduced. Courtship latency, duration of courtship, successful male and duration of copulation were recorded. Vials showing no mating in 1 h were discarded. In series 2, males were handled the same way except that they were stored alone for 1 h between the second and third matings to eliminate any possible effects of postcopulatory preening or inactivity. In series 3, males were stored for 24 h before the third mating and tested with virgin males of the same age.

The number of offspring produced by males mating three times was determined by saving the females from each of the three matings and counting their progeny (Table 1). Fertility is greatly reduced at the third mating when this occurs within minutes of the second mating or when males are rested for an hour. When males are rested for 24 h after the second mating,

their fertility rises to levels seen in once-mated males. The total number of offspring produced by virgin males is slightly less than reported elsewhere². However, the present study used a different wild-type strain.

Table 2 shows the relative mating success of experienced males. The data for the nine replications ($n = 15-20$ per replication) in series 1 were pooled following testing for homogeneity⁵. The numbers shown for series 2 are pooled values for nine replications, and for series 3 eight replications. In series 1 and 2, virgin males had a distinct mating advantage. In series 3, no advantage existed.

Table 2 Relative mating success of virgin and experienced *D. melanogaster* males

Series	% Mating	Virgin	Male chosen Experi- enced	Total	χ^2	P
1	95	115	65	180	13.88	<0.01
2	96	104	74	178	5.02	<0.05
3	96	84	91	175	0.28	NS

Overall clipping $\chi^2 = 0.965$. NS, not significant.

It has been reported that in *Drosophila*, the females 'decide' if and with which male a mating will take place^{3,6}. Obviously, these 'decisions' depend on various sensory cues from the males. It is known that males which spend less time courting have a reduced mating success⁷. However, the disadvantage of the males with reduced fertility in the present study could not be explained by any reduction in observable courtship activity. Experienced males from series 1 and 2, whether successful or not, equalled virgin males in (1) being first to begin courtship, and (2) the proportion of time spent not courting. The average courtship times in series 1 and 2 were 6.01 ± 0.29 min for successful virgin males and 5.97 ± 0.33 min for successful experienced males ($t = 0.024$, NS). The duration of copulation was 20.61 ± 0.91 min and 21 ± 1.06 min for virgin and experienced males, respectively ($t = 0.76$, NS).

Although the virgin and experienced males of series 1 and 2 were of the same age and strain, they are in some way perceived as differing from each other by females. As the standard tests of courtship which would reflect differences in male vigour cannot explain the differences in success of the two types of males, the possibility that they are characterised by different olfactory stimuli⁸, or patterns of organisation of courtship elements⁹ should be investigated. Some stimuli from the males must somehow be influencing the mate selection process of the females. Any mechanism which results in females discriminating the level of fertility of prospective mates would certainly have a selective advantage. Other species should be studied to determine how widespread this phenomenon is.

This work was supported by NIMH grant 30382-01 to T.A.M.

THERESE ANN MARKOW
MAUREEN QUAID
STEVEN KERR

Department of Zoology,
Arizona State University,
Tempe, Arizona 85281

Received 9 February; accepted 8 November.

1. Kvelland, I. *Hereditas* **53**, 281-305 (1965).
2. Lefevre, G. & Johnsson, U. *Genetics* **47**, 1719-1736 (1962).
3. Ehrman, L. *Evolution* **23**, 59-64 (1969).
4. Averhoff, W. & Richardson, R. *Behav. Genet.* **4**, 207-225 (1973).
5. Woolf, C. *Principles of Biometry* (Van Nostrand, New Jersey, 1968).
6. Manning, A. *Anim. Behav.* **15**, 239-250 (1967).
7. Connolly, K., Burnet, B. & Sewell, D. *Evolution* **23**, 548-559 (1969).
8. Leonard, J., Ehrman, L. & Schorsch, M. *Nature* **250**, 261-262 (1974).
9. Bastock, M. & Manning, A. *Behaviour* **8**, 85-111 (1955).

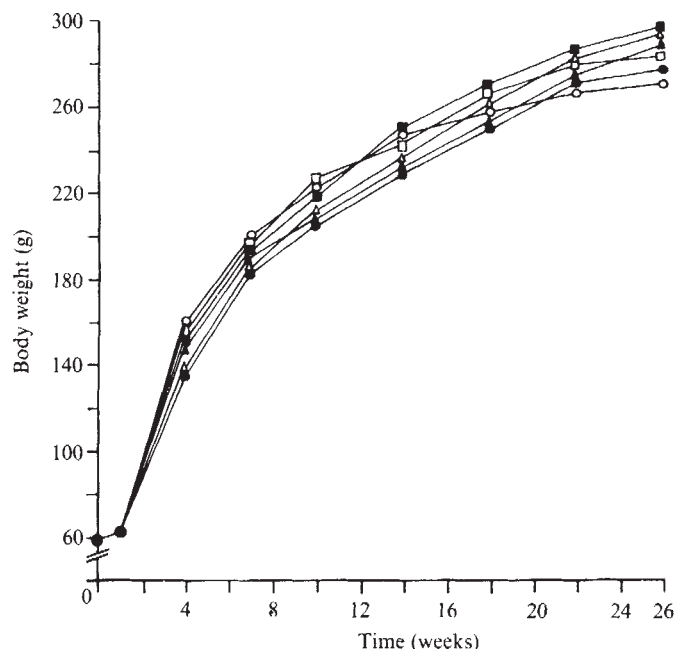
Co-carcinogenic effects of dietary cholesterol in experimental colon cancer

COLORECTAL CANCER is now the most common internal malignancy in the US (ref. 1), but its aetiology is unknown. Western dietary factors are likely to be implicated^{2,3}, and high fat diets correlate well with the geographical incidence of colon cancer⁴⁻⁶. Patients with colon cancer have high levels of faecal bile acids and cholesterol⁷ as do patients with diseases known to be associated with colon cancer, for example, familial polyposis⁸ and ulcerative colitis⁹, and also other high-risk populations¹⁰. It is thought that faecal bile acid and cholesterol metabolites may act as promoters, co-carcinogens or carcinogens in large bowel tumorigenesis¹¹⁻¹³. As cholesterol is the obligatory precursor of the bile acids¹⁴, we have tested the ability of dietary cholesterol to promote the induction of colon cancer by dimethylhydrazine (DMH) in the rat. We report here evidence from an animal study that cholesterol is a potent dietary co-carcinogen.

The experimental method and diets used are detailed in Table 1. The effects of a single dietary variable, cholesterol, was investigated on the induction of colon cancer in two identical groups of rats, both fed on Vivonex, a cholesterol-free, chemically defined liquid diet. One group (V) received unmodified Vivonex, and the other group (VC) received Vivonex with added cholesterol. The third dietary group (S) was included solely for comparison as a carcinogen-injected control group and was fed a standard formula solid diet. By design, all three diets were fibre-free and provided equivalent caloric intakes.

The results were analysed after 56 weeks' observation. During the first 6-month period, animals on all dietary regimens maintained comparable weight gains (Fig. 1). Any observed differences between groups are therefore unlikely to be attributable to the effects of caloric restriction. Comparisons beyond 6 months are inaccurate, because of variations between sick animals in relation to appetite and weight loss due to the induced colon cancers. At 56 weeks, all control animals were alive and well, except for a single rat fed on Vivonex plus cholesterol

Fig. 1 Growth curves showing weight gain of rats on various dietary regimens. Each point represents a mean weight of the 10 rats in each group. Standard deviations have been omitted for clarity. Statistical comparison shows no significant differences between groups at 26 weeks ($F = 0.28$, d.f. = 5,54, $P > 0.10$). Groups: ○, S; ●, V control; △, V; ▲, VC; □, VC control; ■, S control.



which died of pneumonia at 42 weeks with a tumour-free colon. All the remaining deaths occurred in the carcinogen-treated rats and (with two exceptions cited in Table 2) all were attributable to histologically confirmed adenocarcinoma of the colon, or its complications. All animals either died spontaneously (44.4%) or fulfilled one of the recently published¹⁷ objective criteria for killing (55.6%) (ref. 17).

Examination of the results (see Figs 2, 3 and Table 2) shows that DMH-treated animals fed on the cholesterol-free diet Vivonex showed a highly significant increase in both time to tumour presentation ($P < 0.001$) and survival ($P = 0.001$) when compared to carcinogen-injected controls fed on a standard diet. This suggests that Vivonex protected against DMH carcinogenesis. Comparison between the two groups of Vivonex-fed animals (groups V and VC) shows that the addition of cholesterol to Vivonex significantly reduced both the time to tumour presentation ($P < 0.001$) and survival ($P = 0.008$) of group VC, as well as increasing the incidence of metastases ($P = 0.028$). The latter finding was particularly striking (see Table 2) as 67% of group VC animals had abdominal carcinomatosis at necropsy. Distant metastases are uncommon in this model²⁰ and in our experience the incidence rarely exceeds 20%. This suggests that the addition of cholesterol to Vivonex not only abolished the protective effect of Vivonex, producing similar tumour indices in group VC to those of group S, but also enhanced the spread of the induced colonic cancer. These findings are attributable only to the single experimental variable, the provision of cholesterol in the diet.

Table 1 Dietary intake and experimental method

Dietary component	Calculated range of daily intake in each dietary group (mg per adult rat per day)		
	Vivonex* (V)	Vivonex plus cholesterol† (VC)	Standard‡ (S)
Cholesterol	0	6–9§	3–5
Saturated fat	0	0	105–140
Unsaturated fat	65–98	65–98	363–484
Kilocalories per kg	750	750	3580
Consumption per adult rat per day	60–90 ml	60–90 ml	15–20 g
Kilocaloric intake per adult rat per day	45–68	45–68	54–72

Sixty weanling inbred female Wistar rats (A. Tuck) weighing 60–70 g were weaned on to a standard diet (41B), for one week to establish normal colonic function and bacterial colonisation. Thereafter, they were randomly allocated in groups of 20 to one of the three dietary regimens. After a further two weeks on their respective diets, half the animals ($n = 10$) in each group were injected with carcinogen DMH dihydrochloride (97%) (Aldrich) at a dosage of 40 mg per kg body wt per week subcutaneously for 13 weeks and prepared according to the method of Filipe¹⁶. The other half received equivalent volumes of saline preservative to act as controls. They remained on these diets for the rest of their lifetime. Animals were housed in temperature-controlled quarters in subgroups of five in suspended cages which had open-mesh, wire floors designed to prevent coprophagia. They were weighed weekly and inspected daily for signs of illness or colonic disease (for example, rectal bleeding with positive faecal occult blood test, tumour prolapse per rectum, diarrhoea, ascites or colonic obstruction with abdominal distension). When moribund, animals were isolated to prevent cannibalisation, and either died spontaneously or were painlessly killed when they fulfilled objective criteria for death¹⁷. Complete necropsies, including the selection of specimens for histological examination, were performed on all animals. Four tumour indices (see Table 2 for definitions) were measured in each animal during this *in vivo* study. The time to tumour presentation and survival were analysed statistically by the log rank method of Peto *et al.*¹⁸.

* Vivonex was prepared each day as follows: 10 sachets (each providing 300 kcal) were dissolved in 3 l of fresh water, mechanically stirred in a 5 l conical glass flask until fully dissolved, and made up to a final volume of 4 l. It was then divided into 2 equal volumes, to one of which 200 mg cholesterol was added, and both volumes were stirred for 1 h. The liquid diets were then dispensed into standard 540-ml glass feeding bottles and offered to the animals at the same time each day.

† 99.9% free of solid impurities (BDH Chemicals).

‡ Formula Diet 41B, in pellet form (BP Nutrition). This diet was specially prepared to exclude all crude and dietary fibre.

§ This dosage was selected to approximate, on a body weight basis, the daily cholesterol intake of a 70 kg man breakfasting on two fried eggs and two rashers of bacon, that is, 1,000 mg per day¹⁵.

Cholesterol has not yet been shown to be a colon carcinogen, although there is evidence²¹, recently disputed²², that it is sarcomagenic in mice. However, the control animals receiving added dietary cholesterol in our experiment are alive and clinically tumour-free, and the observed effects cannot therefore be ascribed to cholesterol *per se*. As the augmentation of DMH colon tumour induction by cholesterol is within Berenblum's definition of 'co-carcinogenic action'²³, the effects can be attributed to cholesterol acting as a dietary co-carcinogen. We conclude, therefore, that dietary cholesterol facilitates the development, growth and spread of this experimental colon cancer.

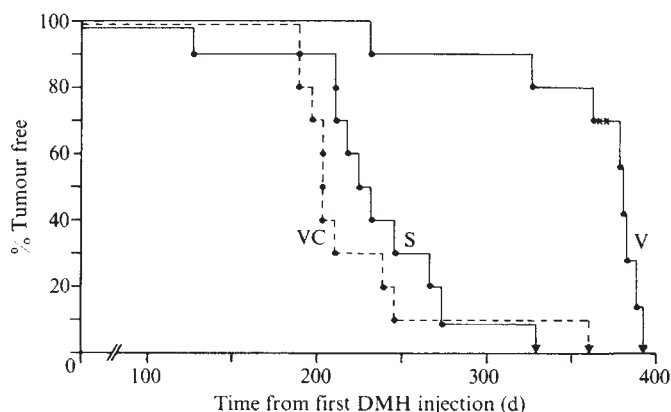


Fig. 2 Life table of proportion of tumour-free animals as a function of time after first injection of DMH. ●, Tumour presentations; x, deaths without presentation. Statistical analysis of time to tumour presentation¹⁸.

Group	Observed nos presenting (O)	Extent of exposure (E)	Ratio (O/E)
V	8	17.70	0.45
VC	10	4.21	2.38
S	10	6.09	1.64

V against VC, $\chi^2 = 14.16$, 1 d.f., $P < 0.001$; VC against S, $\chi^2 = 0.27$, 1 d.f., $P > 0.10$; V against S, $\chi^2 = 13.59$, 1 d.f., $P < 0.001$.

Evidence from other animal studies has shown that high fat diets^{12,24–26} and bile salts^{27,28} promote the effects of several large-bowel carcinogens, and that animals on such high fat diets excrete more faecal neutral sterols and bile acids than do controls¹². However, there is disagreement as to the type of dietary fat responsible for the promotion. Studies comparing the relative role of saturated with unsaturated fats did not detect differences in colon tumour incidence^{25,29}, but a recent report associates an increased tumour yield with a diet high in unsaturated fat and cholesterol²⁶. Such discrepancies are due, in part, to the fact that many experimental diets contain mixtures of fats and variable amounts of non-fat components such as fibre which may exert independent effects²⁵. The use of Vivonex obviates this problem, as it contains neither cholesterol, saturated fats nor dietary fibre and therefore allows carefully controlled dietary alterations. It has been used before in this model by Castleden³⁰, but the experiment was terminated at 24 weeks; the Vivonex diet used was not fibre-free, as the rats had been eating sawdust as well.

Two salient questions are raised by our results. The first relates to the mechanism whereby Vivonex feeding protects against DMH colon carcinogenesis, and the second to the mechanism whereby cholesterol overcomes this protective effect and facilitates tumour development. The effects of both Vivonex and cholesterol on DMH metabolism are unknown. Vivonex has been shown in humans to affect faecal bulk, transit time and flora, as well as reducing the levels of faecal bile acids and cholesterol metabolites³¹, and some or all of these factors may be involved. However, as our evidence demonstrates that

the addition of cholesterol to Vivonex overcomes the protective effect, the simplest explanation is that Vivonex protects because it lacks cholesterol. The protection is clearly incomplete, as Vivonex-fed rats do eventually develop colon cancers. This could be attributed to endogenous synthesis of cholesterol (or its metabolites), or the presence of a small amount of polyunsaturated fat in Vivonex, or a combination of the two. However, it is possible that the induced colon cancers could progress spontaneously in the complete absence of cholesterol, both exogenous and endogenous, due either to random biological accident or to another qualitatively different cause.

The relevance of animal data to the human situation is debatable, but the concepts are useful, particularly when seen against the general framework of multi-stage tumour models. Modern theorists^{32,33} envisage a multi-stage transition from a normal to a cancerous state. In general, the early stages seem to be initiated by mutational events, with subsequent stages proceeding either randomly, by biochemical accident, or at a rate determined by various extrinsic agents. In this context, a parallel may be drawn between animal and human colon carcinogenesis: in the animal model, DMH (a mutagen) presumably affects one or more early stages, and dietary components such as fat and cholesterol one or more later stages. Similarly, whilst the early (presumably mutational) stage in the genesis of human colon

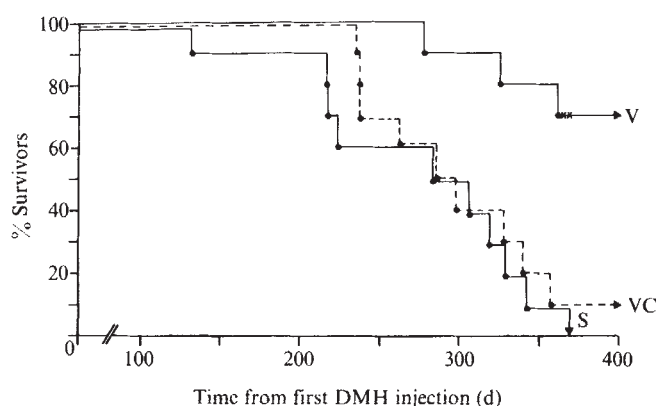


Fig. 3 Life table of proportion of survivors of death from colon cancer as a function of time after first injection of DMH. ●, Deaths from colon cancer; ×, deaths from other causes. Statistical analysis of survival¹⁸:

Group	Observed survival (O)	Expected survival (E)	Ratio (O/E)
V	3	10.98	0.27
VC	9	6.01	1.50
S	10	5.01	2.00

V against VC, $\chi^2 = 6.96$, 1 d.f., $P = 0.008$; VC against S, $\chi^2 = 0.16$, 1 d.f., $P > 0.10$; V against S, $\chi^2 = 10.33$, 1 d.f., $P = 0.001$.

Table 2 Effects of dietary cholesterol on DMH-induced colon cancer in rats, analysis of results 56 weeks after first carcinogen injection

Dietary group (n = 10)	Rat code	Time to tumour presentation* (d)	Survival† (d)	No. of colonic tumours per rat‡	Distant abdominal metastases§
Vivonex (V) (Cholesterol-free)	V1	231	278	1	No
	V3	326	326	2	No
	V7	>362	362	0	No 0/5
	V5	362	362	7	No
	V8	>369	369	1	No
	V10	378			
	V2	380			
	V9	382	Alive at 400 days		
	V6	388			
	V4	392			
Vivonex plus cholesterol (VC)	VC4	189	236	2	Yes
	VC3	189	238	4	No
	VC7	196	238	5	Yes
	VC5	203	263	8	Yes
	VC1	203	286	10	No 6/9
	VC2	203	298	11	Yes
	VC9	210	328	5	Yes
	VC6	238	340	4	Yes
	VC10	245	357	6	No
	VC8	360	Alive at 400 days	-	-
Standard (S)	S8	126	132	1	No
	S1	210	217	3	No
	S5	210	218	7	No
	S4	217	224	4	No
	S10	224	286	4	Yes
	S6	231	306	7	No 2/10
	S2	245	319	2	Yes
	S7	266	329	3	No
	S3	273	342	3	No
	S9	329	369	9	No
All controls (n = 30)	VC	-	297	-	No
	Control	2			

* Defined as time in days from first carcinogen injection to the first sign of colonic disease. Statistical analysis, see legend to Fig. 2.

† All deaths attributable to colon cancer, except single control death. Animals V7 and V8 were killed for humane reasons after they had developed large primary malignant renal tumours. Neither had colonic carcinoma at histology. Statistical analysis, see legend to Fig. 3.

‡ Macroscopic tumours counted at necropsy. At least one tumour per animal was histologically confirmed as a primary adenocarcinoma of the colon. Statistical analysis, no significant differences between groups using Mann-Whitney test¹⁹.

§ Distant abdominal metastases (to regional nodes, omentum, peritoneum and liver), confirmed histologically. Statistical analysis, the incidence of metastases in group VC was significantly greater than that in group V ($P = 0.028$) using the one-tailed Fisher exact test¹⁸.

cancer is unknown, the disease does seem to involve multi-stage processes³⁴, and an interpretation of the data implicating dietary fat and cholesterol in man would be that these dietary components are analogous rate-determining extrinsic agents affecting the later stages.

Although it is possible, in both animals and man, that the different stages could proceed in the absence of dietary fat or cholesterol, the experimental and epidemiological evidence is consistent with the interpretation that these dietary components are associated with an apparent acceleration of the processes involved. Our observations corroborate this concept, and specifically suggest that dietary cholesterol is a factor determining the rate of development of experimental colon cancer. If this applies to man, as the metabolic evidence would suggest, then human colon cancer may be a preventable disease.

We thank Mrs F. Briscoe and Eaton Laboratories for supplying and sharing the cost of Vivonex, Mr P. Luther for technical assistance, the animal house staff, Miss Susan White of the MRC Statistical Research and Services Unit and Miss Diana Wilson for typing the manuscript. J.P.C. is a Rhodes Scholar and acknowledges the support of the Rhodes Trust, Oxford.

J. P. CRUSE
M. R. LEWIN
G. P. FERULANO
C. G. CLARK

*Surgical Unit,
University College Hospital Medical School,
University Street,
London, WC1, UK*

Received 14 August; accepted 9 November 1978.

- Cutler, S. J. & Young, J. I. in *Persons at High Risk of Cancer* (ed. Fraumeni, J. F. Jr) 307-342 (Academic, New York, 1975).
- Wynder, E. L. & Shigematsu, T. *Cancer* **20**, 1520-1561 (1967).
- Burkitt, D. P. *Cancer* **28**, 3-13 (1971).
- Weisburger, J. H., Reddy, B. S. & Wynder, E. L. *Cancer* **40**, 2414-2420 (1977).
- Wynder, E. L., Kajitani, T., Ishikawa, S., Dodo, H. & Takano, A. *Cancer* **23**, 1210-1220 (1969).
- Drasar, B. S. & Irving, D. *Br. J. Cancer* **27**, 167-172 (1973).
- Hill, M. J. *et al. Lancet* **i**, 535-539 (1975).
- Reddy, B. S., Mastromarino, A., Gustavson, C., Lipkin, M. & Wynder, E. L. *Cancer* **38**, 1694-1698 (1976).
- Reddy, B. S., Martin, C. W. & Wynder, E. L. *Cancer Res.* **37**, 1697-1701 (1977).
- Hill, M. J. *et al. Lancet* **i**, 95-100 (1971).
- Hill, M. J. *Digestion* **11**, 289-306 (1974).
- Reddy, B. S., Weisburger, J. H. & Wynder, E. L. *J. natn. Cancer Inst.* **52**, 507-511 (1974).

13. Wynder, E. L. & Reddy, B. S. *Cancer* **40**, 2565-2571 (1977).
14. Nair, P. P. & Kritchevsky, D. (eds) *The Bile Acids* Vols I & II (Plenum, New York, 1971).
15. Paul, A. A. & Southgate, D. A. T. *The Composition of Foods*, 4th edn (eds McCance, R. A. & Widdowson, E. M.) (MRC Special Rep. no. 297, HMSO, London, 1978).
16. Filipe, M. I. *Br. J. Cancer* **32**, 60-76 (1975).
17. Cruse, J. P., Lewin, M. R. & Clark, C. G. *Br. J. Cancer* **37**, 639-643 (1978).
18. Peto, R. *et al. Br. J. Cancer* **35**, 1-39 (1977).
19. Siegel, S. *Nonparametric Statistics for the Behavioural Sciences* (McGraw-Hill, New York, 1956).
20. Rogers, A. E. & Newberne, P. M. *Cancer Res.* **35**, 3427-3431 (1975).
21. Heiger, I. *Br. med. Bull.* **14**, 159-160 (1958).
22. Hill, M. J. *Crit. Rev. Tox.* **4**, 32-82 (1975).
23. Berenblum, I. *Prog. exp. Tumor Res.* **11**, 21-30 (1969).
24. Nigro, N. D., Singh, D. V., Campbell, R. L. & Pak, M. S. *J. natn. Cancer Inst.* **54**, 439-442 (1975).
25. Wilson, R. B., Hutcheson, D. P. & Wideman, L. *Am. J. clin. Nutr.* **30**, 176-181 (1977).
26. Broitman, S. A., Vitale, J. J., Vavrousek-Jakuba, E. & Gottlieb, L. S. *Cancer* **40**, 2455-2463 (1977).
27. Nigro, N. D., Bhadrachari, N. & Chomchai, C. *Dis. Colon Rectum* **16**, 438-443 (1973).
28. Narisawa, T., Magadia, N. E., Weisburger, J. H. & Wynder, E. L. *J. natn. Cancer Inst.* **53**, 1093-1097 (1974).
29. Reddy, B. S., Narisawa, T., Vukusich, D., Weisburger, J. H. & Wynder, E. L. *Proc. Soc. exp. Biol. Med.* **151**, 237-239 (1976).
30. Castleden, W. M. *Br. J. Cancer* **35**, 491-495 (1977).
31. Crowther, J. S., Drasar, B. S., Goddard, P., Hill, M. J. & Johnson, K. *Gut* **14**, 790-793 (1973).
32. Peto, R. in *Origins of Human Cancer* 1403-1428 (Cold Spring Harbour, New York, 1977).
33. Berenblum, I. *J. natn. Cancer Inst.* **60**, 723-726 (1978).
34. Hill, M. J., Morson, B. C. & Bussey, H. J. R. *Lancet* **i**, 245-247 (1978).

Protease inhibitors suppress radiation-induced malignant transformation *in vitro*

LITTLE is known about the mechanisms of carcinogenesis. The fact that most carcinogens are mutagenic has led to speculation that the primary step in cancer induction may be mutational^{1,2}; there is evidence from both *in vivo*² and *in vitro*³ studies that a strong correlation exists between the mutagenicity and carcinogenicity of an agent. Mutagenic and carcinogenic agents, both physical and chemical, also produce similar kinds of DNA damage and repair⁴. Radiation-induced mutagenesis in some bacterial cells requires an error-prone DNA repair system^{5,6}, and there is now some evidence that error-prone DNA repair may be involved in the malignant transformation of cells by radiation^{7,8}. Protease inhibitors have been shown to suppress

specifically both error-prone repair and mutagenesis in bacterial cells^{9,10}, as well as to inhibit carcinogenesis *in vivo*^{11,12}. We report here that the protease inhibitors antipain and leupeptin will suppress radiation-induced transformation *in vitro* as well as inhibit two-stage transformation *in vitro* using radiation and the promoting agent, 12-O-tetradecanoyl-phorbol-13-acetate (TPA).

We used a C3H mouse-embryo-derived cell line (10T $\frac{1}{2}$ clone 8) isolated and characterised by Reznikoff *et al.*^{13,14}, and adapted for studies of radiation transformation^{7,15,16}. In some experiments, a BALB/3T3 mouse line designated A31-11 was used. Concentrations of the protease inhibitors used were nontoxic, either with or without radiation exposure, and have been shown not to affect growth, RNA or protein synthesis in bacterial⁹ or mammalian cells¹⁸⁻²⁰.

The influence of antipain and leupeptin on X-ray induced transformation is shown in Table 1. In these experiments, the inhibitors were present for the entire post-irradiation expression period. As can be seen, antipain not only suppressed X-ray-induced transformation but also completely abolished transformation enhanced by TPA. Similarly, leupeptin suppressed radiation transformation enhanced by TPA, although this protease inhibitor did not seem to be as effective as antipain. To determine whether these protease inhibitors were effective in suppressing radiation transformation when present for only a short post-irradiation interval, experiments were carried out with antipain present only during the 5-d or 1-d period following radiation exposure. Antipain at 1 mM (600 $\mu\text{g ml}^{-1}$) reduced transformation from 2.0×10^{-3} to $<3.2 \times 10^{-4}$ when present for 5 d and to 5.5×10^{-4} when present for only 1 d following 600 rads.

A preliminary experiment was carried out with A31-11 mouse BALB/3T3 cells to determine whether the suppressive effect of antipain was cell line-specific. These results are shown in Table 2. Again, incubation with antipain (50 $\mu\text{g ml}^{-1}$) throughout the expression period completely suppressed transformation induced by 100 or 400 rad of X rays plus TPA exposure.

Endogenous proteases have been implicated previously in the mechanism of tumour promotion²¹, and protease inhibitors have been shown to suppress two-stage carcinogenesis *in*

Table 1 Effects of 6 weeks post-irradiation exposure to antipain and leupeptin on X-ray transformation of 10T $\frac{1}{2}$ cells with and without promotion by TPA

Treatment	Expt. no.	PE %	Total no. cells at risk	Total no. transformed foci*	Transformation frequency per viable cell ($\times 10^{-4}$)*
Controls (no treatment)	1	32.5	3,120	0	<3.2
	2	29.3	4,395	0	<2.3
Antipain alone	1	32.2	3,088	0	<3.2
	2	32.3	4,845	0	<2.1
600 rads	1	3.0	2,160	5	23.1
	2	2.4	7,200	11	15.3
600 rads + antipain	1	3.4	3,264	1	3.1
	2	2.7	3,120	0	<3.2
600 rads + leupeptin	1	3.8	3,192	1	3.1
	2	2.9	3,780	2	5.2
400 rads	1	6.3	3,402	3	8.8
	2	4.9	7,200	4	5.5
400 rads + TPA	1	7.0	2,940	9	30.6
	2	5.7	7,140	22	30.8
400 rads + TPA + antipain	1	7.4	1,332	0	<7.5
	2	7.7	4,620	0	<2.2
400 rads + TPA + leupeptin	1	10.6	5,088	2	3.9

Details of experimental techniques for radiation transformation experiments using 10T $\frac{1}{2}$ cells have been described elsewhere^{7,15,16}. Stock cultures were maintained in 60 mm Petri dishes and were passed by subculturing at a 1:20 dilution every 7 d. The cells used were in passages 9-14. They were grown in a humidified 5% CO₂ atmosphere at 37 °C in Eagle's basal medium supplemented with 10% heat-inactivated fetal calf serum and antibiotics. Cells were irradiated 24 h after seeding, and incubated in fresh medium containing TPA and/or protease inhibitors immediately afterwards. Antipain and leupeptin added at 50 $\mu\text{g ml}^{-1}$; TPA at 0.1 $\mu\text{g ml}^{-1}$.

* Types 2 and 3 foci were scored as transformants, as described by Terzaghi and J.B.L.¹⁵.

Table 2 Effect of antipain on oncogenic transformation induced in 3T3 cells by X rays plus TPA

Treatment	Total no. cells at risk	Total no. transformed foci	Transformation frequency ($\times 10^{-4}$)
400 rads	60,500	66	10.9
400 rads + antipain	43,450	14	3.2
400 rads + TPA	52,800	91	17.2
400 rads + TPA + antipain	50,050	0	<0.2
100 rads + TPA	34,400	29	8.4
100 rads + TPA + antipain	33,500	0	<0.3

Stock cultures were passaged as for $10T\frac{1}{2}$ cells. Cells were seeded 24 h before irradiation at densities which yielded about 4,000 viable cells per 100 mm Petri dish. They were grown in Eagle's minimal essential medium supplemented with 10% heat-inactivated fetal calf serum and antibiotics. Transformed foci were scored at 4 weeks¹⁷. Antipain was added at $50 \mu\text{g ml}^{-1}$ (0.08 mM) and TPA at $0.1 \mu\text{g ml}^{-1}$.

in vivo^{11,12}. It is possible, therefore, that the suppression of radiation transformation by these protease inhibitors is due entirely to an effect on the promotion stage of transformation (or carcinogenesis). This hypothesis is supported by the observation that almost complete suppression occurred in the presence of TPA. However, the fact that antipain was also effective in inhibiting transformation at early stages (within the first 1 or 5 d of irradiation) argues against this hypothesis, and suggests that an effect on DNA repair processes might be important in its action. TPA is ineffective in enhancing X-ray transformation unless it is present for at least 14 d or the entire proliferative phase of expression after irradiation (unpublished data).

Mutagenesis in some bacterial cells requires an error-prone DNA repair ('SOS repair') function^{5,6}. It has been postulated that the induction of mutations in mammalian cells may also be associated with an error-prone DNA repair process^{22,23}. There is some evidence that DNA repair mechanisms operating within the first few hours of radiation exposure may be involved in the induction of malignant transformation by X irradiation in $10T\frac{1}{2}$ cells^{7,8}. The theory that proteases have the ability to cleave suppressors of error-prone repair in bacteria and thus induce this type of repair^{5,6,9} suggests that protease inhibitors might inhibit malignant transformation *in vitro* by inhibiting error-prone repair. Thus, it is possible that the action of protease inhibitors on a DNA repair process may be a mechanism for our observed suppression of radiation transformation. Further investigation of this hypothesis will require a careful examination of the post-irradiation intervals during which antipain has its greatest effect on transformation, as well as its effect on specific molecular repair processes during these times.

In any event, the present results offer *in vitro* confirmation for the suppressive effect of protease inhibitors on carcinogenesis *in vivo*. They suggest that a further examination of the mechanisms of action of these agents may offer valuable insights into mechanisms of carcinogenesis.

We thank Dr Miroslav Radman for suggesting these experiments to us, Mr Gary Murphy for expert technical assistance, Dr T. Kakunaga for the A31-11 line and Dr Walter Troll and the US-Japan Cooperative Cancer Research Program for providing the antipain and leupeptin used in this study. This research was supported by NIH grants CA-22704 and ES-00002.

ANN R. KENNEDY
JOHN B. LITTLE

Laboratory of Radiobiology,
Department of Physiology,
Harvard University,
School of Public Health,
Boston, Massachusetts 02115

Received 14 August; accepted 6 November 1978.

- McCann, J., Choi, E., Yamasaki, E. & Ames, B. N. *Proc. natn. Acad. Sci. U.S.A.* **72**, 5135-5139 (1975).
- Ames, B. *Science* **191**, 241-244 (1976).
- Huberman, E. & Sachs, L. *Int. J. Cancer* **13**, 326-333 (1974).

- Regan, J. D. & Setlow, R. B. *Cancer Res.* **34**, 3318-3325 (1974).
- Witkin, E. M. *Bact. Rev.* **40**, 869-907 (1976).
- Radman, M., Villani, G., Boiteux, S., Defais, M. & Caillet-Fauquet, P. in *Origins of Human Cancer* (eds Hiatt, H. H., Watson, J. D. & Winston, J. A.) 903-922 (Cold Spring Harbor Laboratories, New York, 1977).
- Terzaghi, M. & Little, J. B. *Nature* **253**, 548-549 (1975).
- Little, J. B. in *Origins of Human Cancer* (eds Hiatt, H. H., Watson, J. D. & Winston, J. A.) 923-939 (Cold Spring Harbor Laboratories, New York, 1977).
- Meyn, M. S., Rossman, T. & Troll, W. *Proc. natn. Acad. Sci. U.S.A.* **74**, 1152-1156 (1977).
- Umezawa, K., Matsushima, T. & Sugimura, T. *Japan Acad. Ser. B* **53**, 30-33 (1977).
- Troll, W. A., Klassen, A. & Janoff, A. *Science* **169**, 1211-1213 (1970).
- Hozumi, M., Ogawa, M., Sugimura, T., Takeuchi, T. & Umezawa, H. *Cancer Res.* **32**, 1725-1728 (1972).
- Reznikoff, C. A., Brankow, D. W. & Heidelberger, C. *Cancer Res.* **33**, 3231-3238 (1973).
- Reznikoff, C. A., Bertram, J. S., Brankow, D. W. & Heidelberger, C. *Cancer Res.* **33**, 3239-3249 (1973).
- Terzaghi, M. & Little, J. B. *Cancer Res.* **36**, 1367-1374 (1976).
- Kennedy, A. R., Mondal, S., Heidelberger, C. & Little, J. B. *Cancer Res.* **38**, 439-443 (1978).
- Kakunaga, T. *Int. J. Cancer* **12**, 463-473 (1973).
- Weber, M. J., Hale, A. H. & Roll, D. E. in *Proteases and Biological Control* (eds Reich, E., Rifkin, D. B. & Shaw, E.) 915-930 (Cold Spring Harbor Laboratories, New York, 1975).
- Hynes, R. O., Wyke, J. A., Bye, J. M., Humphries, K. C. & Pearlstein, E. S. in *Proteases and Biological Control* (eds Reich, E., Rifkin, D. B. & Shaw, E.) 931-944 (Cold Spring Harbor Laboratories, New York, 1975).
- Blumberg, P. M. & Robbins, P. W. in *Proteases and Biological Control* (eds Reich, E., Rifkin, D. B. & Shaw, E.) 945-956 (Cold Spring Harbor Laboratories, New York, 1975).
- Troll, W. in *Fundamentals in Cancer Prevention* (eds Magee, P. N. et al.) 41-55 (University Park Press, Baltimore, 1976).
- Fox, M. *Mutat. Res.* **24**, 187-204 (1974).
- Trosko, J. E. & Chu, E. H. Y. *Chem. biol. Interact.* **6**, 317-332 (1973).

Immunological properties of the surface of parasitic nematodes

THE surface of parasites seems to be especially important in the intricate relationship which is known to exist between these organisms and their hosts. This is particularly clear in the case of African trypanosomes, which are able to vary the antigenic glycoproteins of their surface sequentially during infection¹. It is also known that the surface of schistosomes changes during infection in parallel with alterations in the susceptibility of this trematode to the action of immunity². By comparison with these and other parasites, little is known of the properties of the surface of nematodes nor of its role in the host-parasite interaction. Both free-living and parasitic nematodes are bounded by a non-cellular cuticular layer which is probably living because it may contain, for example, enzymes and haemoglobin³. The cuticle, which is shed and replaced as the nematodes pass through the various moults in their life cycle, consists of fibrous collagen-like proteins, thought unlikely by some authors to be highly immunogenic in the host⁴. However, we now report that

Table 1 Viability of *T. spiralis* newborn larvae cultured *in vitro* for 30 h with peritoneal cells and immune serum

Source of serum	Source of peritoneal cells	% Parasites dead*
Normal rat	Normal rats	2.2
	Infected rats	0.5
Normal rat, heat inactivated	Normal rats	5.8
	Infected rats	2.1
Immune rat	Normal rats	93.7
	Infected rats	98.6
Immune rat, heat inactivated	Normal rats	66.6
	Infected rats	36.1

Immune serum was taken at either the 18th or 24th day after a primary infection with 4,000 infective larvae. Inactivation was achieved by maintaining serum at 56°C for 1 h. Cells from infected rats were taken from the peritoneal cavity of rats 24 d after a primary infection and prepared as described in Fig. 1. This suspension contained 35-45% eosinophils, 5% mast cells and 50-60% mononuclear cells.

* 150 Parasites per replicate were observed with at least four replicates in each group. The parasites were considered to be dead when they were completely immobilised and had visible signs of deterioration in their internal organs.

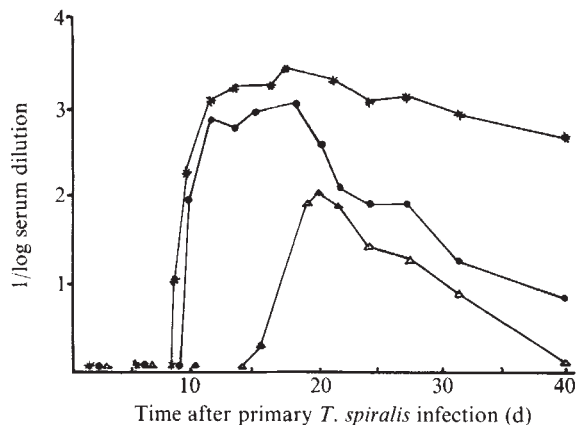


Fig. 1 Serum studies in August rats infected with 4,000 infective larvae of *T. spiralis*. The time after infection to appearance of antibodies mediating *in vitro* adherence of a population of peritoneal cells to newborn larvae (Δ), infective larvae (*) and adult parasites ($\bullet$). Sera obtained at various times after infection were inactivated by heating at 56 °C for 1 h. The peritoneal cells were incubated on a plastic surface for 2 h and the non-adherent supernatant cells used in this assay. The parasites were prepared by standard methods¹³. 0.3-ml cultures were set up in flat-based microtitre trays^{5,6} to provide a cell:parasite ratio of 2,500:1 for newborn larvae, 4,000:1 for infective larvae and 8,000:1 for adult worms. These cultures were incubated at 37 °C for 90 min before assessing the cell adherence to the surface of the parasites by microscopic observation. The end point of the assay was taken as the serum dilution in which 50% of the parasite's surface area was covered with cells. The adherent cells were identified by staining with Giemsa and with chromotrope R, the distribution being approximately 95% eosinophils and 5% mast cells. The points indicated represent the means of at least three experiments, each carried out in triplicate. The s.d. value at each point did not exceed 0.25 except with infective larvae on days 18 and 20, where the s.d.'s were 0.7 and 0.4, respectively. With newborn larvae the values were less than 0.25 except on days 18 and 20 where they were 0.4 and 0.6.

the serum of animals infected with the nematodes *Trichinella spiralis* and *Nippostrongylus brasiliensis* contain antibodies to nematode cuticle, each specific to the surface of a particular stage in the life cycle of the parasites and capable of mediating attack by inflammatory cells against the nematode surface.

The presence in the serum of rats of factors interacting with the surface of nematodes was assessed by studying the adherence to the worms *in vitro* of peritoneal cell populations obtained from normal August rats, as described previously^{5,6}. The cell populations used in this study were obtained by incubating peritoneal cells on a plastic surface for 2 h at 37 °C and collecting the non-adherent cells which consisted of 60–80% eosinophils, 15–20% mast cells and 5–20% mononuclear cells. Initial studies showed that even in serum freshly obtained from uninfected rats, cells adhered to the surface of some but not all stages in the life cycle of the nematodes studied. In dilutions of fresh normal serum, cells adhered to infective larvae of *T. spiralis* and *N. brasiliensis* to a titre of 1/300 and to adult worms of both species to a dilution of 1/60, but did not adhere to the newborn larvae of *T. spiralis*, which is the stage shed by the viviparous mature females of this species. Adherence was markedly reduced after the normal serum had been inactivated by heating at 56 °C for 1 h or had been treated with zymosan (5 mg per ml serum; Zymosan A, Sigma)⁷.

These results show that the surface of infective larvae and adult worms activate complement but that the cuticle of newborn *T. spiralis* larvae does not have this property. However, heat-stable factors which cause adherence of the eosinophil-rich cell population to all stages of these parasites occur in the serum of infected animals. Figure 1 shows the occurrence of these factors in the serum of rats infected orally with 4,000 *T. spiralis*

infective larvae. The appearance and maximum titre of these factors were different for each stage of the parasite, suggesting the presence of antibodies specific for each life-cycle stage. Adsorption studies confirmed this. When serum taken from rats 14 d after infection was absorbed with either infective larvae or adult *T. spiralis*, the serum factor mediating cellular adherence to the homologous but not heterologous life-cycle stage was removed (Fig. 2). Similar adsorption and time-course studies have shown that the serum factor mediating cell adherence to *N. brasiliensis* infective larvae differs from that mediating adherence to adult *N. brasiliensis*⁸. Because of their intense specificity for each stage, we assume that these serum factors are antibodies.

Although antibodies to the surface of nematodes have frequently been described using such methods as immunofluorescence^{9,10} or cell adherence^{6,11,12}, their significance in terms of the nature of the parasite surface and its role in the host-parasite reaction has not been assessed. Further studies using serum and cells from normal or infected rats showed that the adherence of cells in the presence of serum from infected animals may have a lethal effect on these parasites at least *in vitro*. *T. spiralis* newborn larvae were killed after incubation for 25–35 h in the presence of serum from either infected or uninfected donors (Table 1). Fresh normal serum did not affect the viability of the parasites even in the presence of cells. However, the lethal effect of cells and the serum factors from infected animals was greatly enhanced when freshly collected, rather than heat-inactivated, serum was used. Thus, complement seems to affect newborn larvae of *T. spiralis* only in the presence of antibody. Whether or not cells must adhere to produce damage has not yet been clearly determined; however, dead or dying larvae always had adherent cells, generally eosinophils. Preliminary experiments with purified cell suspensions implicate this cell type as an effector cell in causing damage to newborn larvae in this *in vitro* model. The poorer ability of the cells from infected animals to damage larvae may therefore be due to the lower proportion of eosinophil leukocytes in these cell suspensions. A percentage of infective larvae of *T. spiralis* and *N. brasiliensis* can also be damaged by incubation in fresh serum taken from infected animals together with peritoneal cells, but we have not yet fully defined the parameters of this reaction.

This study has revealed three features of the surface of nematode parasites. The surface of infective larvae and adults of *T. spiralis* and *N. brasiliensis* may activate complement, but the surface of newborn larvae does not have this property. During

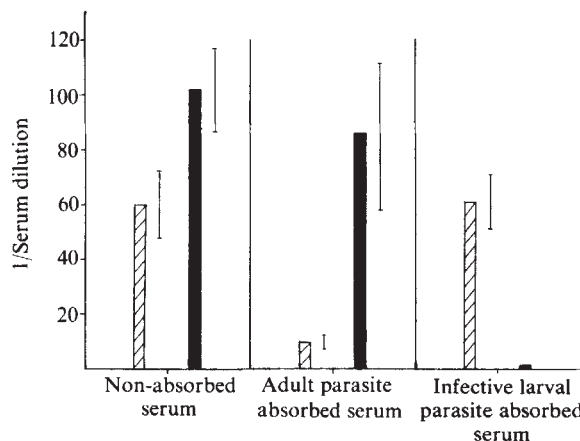


Fig. 2 Adsorption studies to show the specificity of the serum factors mediating cell adherence to *T. spiralis*. Serum obtained 14 d after infection with 4,000 *T. spiralis* infective larvae was assayed for its ability to mediate cell adherence before and after adsorption with adults ($\square$) or infective ($\blacksquare$) larvae. An equal volume of freshly prepared parasites was added to a 1 in 3 dilution of the previously heat-inactivated serum and incubated at 37 °C for 90 min. The supernatant was then assayed as described in Fig. 1; s.d. values are expressed as bar lines.

infection, each stage in the life cycle of *T. spiralis* or *N. brasiliensis* stimulates antibodies which are specific for the surface of that stage only. Finally, in the presence of these antibodies, the parasites are vulnerable to attack and may be killed by peritoneal cell populations enriched in eosinophils, especially in freshly collected serum. Variations in the interaction between the host and different parasitic stages of an infecting nematode are likely to be due in part to differences in the interaction of these stages with complement, antibody and inflammatory cells.

C. D. MACKENZIE
PATRICIA M. PRESTON
BRIDGET M. OGILVIE

Division of Parasitology,
National Institute for Medical Research,
Mill Hill, London NW7, UK

Received 4 August; accepted 14 November 1978.

1. Cross, G. A. M. *Proc. R. Soc. Lond. B* **202**, 55–72 (1978).
2. Smithers, S. R., McLaren, D. J. & Ramalho-Pinto, F. J. *Am. J. trop. Med. Hyg.* **26**, 11–19 (1977).
3. Lee, D. L. *Adv. Parasit.* **4**, 187–254 (1966).
4. Martinez-Palomo, A. *J. Parasit.* **64**, 127–136 (1978).
5. Mackenzie, C. D., Ramalho-Pinto, F. J., McLaren, D. J. & Smithers, S. R. *Clin. exp. Immun.* **30**, 97–104 (1977).
6. McLaren, D. J., Mackenzie, C. D. & Ramalho-Pinto, F. J. *Clin. exp. Immun.* **30**, 105–118 (1977).
7. Fine, D. P., Marney, S. R., Colley, D. G., Sergeant, J. S. & Des Pres, R. M. *J. Immun.* **109**, 807–809 (1972).
8. Ogilvie, B. M., Mackenzie, C. D. & Love, R. J. *Am. J. trop. Med. Hyg.* **26**, 61–67 (1977).
9. Sadun, E. H. *Expl Parasit.* **13**, 72–82 (1963).
10. Ponnudurai, T., Denham, D. A., Nelson, G. S. & Rogers, R. J. *Helminth.* **48**, 107–111 (1975).
11. Soulsby, E. J. L. *Ann. N.Y. Acad. Sci.* **113**, 492–509 (1963).
12. Higashi, G. I. & Chowdhury, A. B. *Immunology* **19**, 65–83 (1970).
13. Dennis, D. T., Despommier, D. & Davis, N. J. *Parasit.* **56**, 974–977 (1970).

Interferon inhibits the release of plasminogen activator from SV3T3 cells

INTERFERON, as well as inducing an antiviral state, can modify certain other cell characteristics. For example, it has been shown to inhibit cell proliferation^{1–4} and release of C-type virus^{5,6}, and to enhance the expression of H-2 antigens^{7,8} and the amount of glycoprotein on the cell surface⁹. These effects might be explained by decreased release of cell-surface components into the medium. Similarly, the expression of a glycoprotein on the cell surface has been related to the growth state of cells¹⁰, which can be explained in part by differential release^{11,12}. We have followed up this suggestion by examining the effect of interferon on the release of plasminogen activator (PA) by a line of simian virus 40 transformed Swiss 3T3 (SV3T3) cells. Our results, reported here, indicate that interferon inhibits the release of PA by SV3T3 cells.

First, to establish the sensitivity of SV3T3 cells to the anticellular activity of interferon, we grew cells in the presence of 50, 500, or 5,000 units ml⁻¹ of a crude preparation of interferon¹³. Cell numbers (day 3) were reduced by 15, 40, and 50%, respectively, by these doses. These results were confirmed using the semi-purified mouse L cell interferon (3 × 10⁷ NIH units per mg protein, obtained from Dr K. Paucker¹⁴) which was used for the rest of the study. Similar observations have been reported for several other cell types which respond to the anticellular properties of interferon^{1–4}.

The release of PA by SV3T3 cells is independent of cell density¹⁵. To determine whether interferon alters this pattern, we grew SV3T3 cells in the presence of interferon for 24 h and measured the PA released into a fresh serum-free harvest fluid (HF) during the next 10–12 h. After collection of the HF, the cells were lysed with Triton X-100, and lysate PA content also determined. Treatment with 50, 100, or 500 NIH reference units interferon ml⁻¹ resulted in 37, 54, and 58% reductions, respectively, in the amount of PA released into the HF (significant by Students' *t* test, *P* < 0.005) (Table 1a). In contrast, cell lysate PA activity was not significantly reduced. In

Table 1 The effects of interferon and cycloheximide on PA activity and protein synthesis

	% of control units PA†		³ H-amino acid incorporation‡ (c.p.m. mg protein ⁻¹)
	HF	Lysate	
a. Interferon*			
(Units ml ⁻¹)			
0	100 ± 4	100 ± 2	5,994 ± 177
50	63 ± 5	97 ± 3	6,308 ± 171
100	46 ± 4	88 ± 4	6,395 ± 130
500	42 ± 8	106 ± 6	5,646 ± 166
b. Cycloheximide§			
(µg ml ⁻¹)			
0	100 ± 4	100 ± 2	3,286 ± 317
0.1	64 ± 5	96 ± 3	1,934 ± 35
0.2	31 ± 1	64 ± 4	1,044 ± 23
0.5	19 ± 1	40 ± 1	650 ± 26

*Interferon was assayed in 3T3 cells by a neutral red dye uptake method patterned after that of Finter²⁶ using vesicular stomatitis virus as the challenge virus. Units are expressed as reference units based on NIH reference standard G-002-904-51.

†Cells were plated at 5 × 10⁴ cells per 35-mm plastic petri dish (Falcon) in Dulbecco's modification of Eagle's minimal essential medium (Gibco) supplemented with fetal calf serum to a final concentration of 5%. Cells were incubated at 37 °C in a humidified atmosphere of 5% CO₂ in air for 18 h before addition of interferon with medium change. Cell counts were performed in duplicate on cells resuspended by trypsinisation. Cells were counted using a haemocytometer. For preparation of PA samples, cells were grown to a density of 2 × 10⁵–3 × 10⁵ per dish, switched to serum-free medium ± interferon, and the HF collected after an additional 10–12 h of incubation at 37 °C. Triton X-100 cell lysates were prepared, and HF and lysate samples assayed for PA activity using ¹²⁵I-fibrin as the substrate as previously described²⁷, with slight modifications. The assay mixture contained: 0.1% Triton X-100 in 0.1 M Tris-HCl, pH 8.1, 5–10 µg of purified human plasminogen²⁸, and HF or lysate sample in 1.0 ml final volume. One unit of PA is the amount of enzyme generating enough plasmin to catalyse the hydrolysis of 10% of the total fibrin substrate in 4 h at 37 °C. Plasminogen blanks were subtracted before calculation of the number of units, and data were normalised for cell number or for total protein in Triton X-100 lysates, as determined by the procedure of Lowry *et al.*²⁹. Each value is the mean of data from three separate experiments. Each experiment was carried out in duplicate, and each resulting sample assayed two or three times. Data are expressed as % of control values to facilitate comparisons of data derived from separate experiments. Data are shown ± s.e.m.

‡³H-amino acid incorporation was measured at the end of the HF collection using dishes treated in parallel with those used to measure PA. For labelling, cells were washed twice with 37 °C Hank's balanced salt solution (HBSS) containing appropriate doses of interferon or cycloheximide. Cells were labelled for 1 h with 2.5 µCi ³H-L-amino acid mixture (New England Nuclear) in 1.0 ml of HBSS containing interferon or cycloheximide. After 1 h the medium was removed, and the cells washed three times with warm HBSS, extracted three times with cold 5% trichloroacetic acid (TCA), and solubilised in 0.6 ml of 0.5 M NaOH. Fetal calf serum (20 µl) was added to duplicate 0.2 ml portions of the solution to serve as unlabelled carrier protein, and the samples neutralised with 4 M HCl. Neutralised samples were precipitated with 2.0 ml of cold TCA and stored at 4 °C overnight. TCA precipitates were collected on Whatman GF/C filters and the filters washed exhaustively with cold 5% TCA, dried, and counted in 10 ml of Omnifluor (NEN) toluene. Data were normalised for protein content of the NaOH solutions as described above and represent mean values ± s.e.m. of duplicate determinations from each of two cultures.

§Cycloheximide (Sigma) was added at the beginning of HF preparation for PA studies.

two other experiments, up to 1,000 units interferon ml⁻¹ did not inhibit PA release any further. Interferon doses higher than 200 units ml⁻¹ have resulted in significant decreases in cell lysate PA in some experiments. Direct addition of interferon (up to 500 units ml⁻¹) to the PA assay, or preincubation of control PA samples with interferon for 1 h at 37 °C, did not decrease the assayable PA.

In spite of the relative purity of the interferon used, the participation of residual impurities was considered. Pretreat-

ment of interferon (100 units ml⁻¹, final concentration) for 1 h at 37°C with a 1:200 dilution of rabbit anti-mouse interferon serum (titre 4,000, Reference Reagents Branch, NIH) inactivated its antiviral activity, as well as its inhibition of PA release (inhibited by only 13%). In parallel experiments, interferon treated with preimmune serum, or with diluent, retained activity, inhibiting PA release by 46 and 48% respectively. Column void material from an anti-interferon globulin affinity column used for interferon purification¹⁴ did not inhibit PA release even when used at a pre-column equivalent of 10,000 units of interferon ml⁻¹. Heating interferon at 60°C for 1 h destroyed its effect on PA release, but dialysis at 4°C for 24 h did not. These properties of mouse interferon are consistent with those reported by Gresser *et al.*¹, and indicate that the observed inhibition of PA release is due to interferon itself.

Interferon does not cause accumulation of PA in SV3T3 cells (Table 1a), and at high doses may decrease cellular PA (data not shown). Further, high doses of interferon inhibit protein synthesis in mouse L cells¹⁶. Thus, the possibility that the effect of interferon on PA release might be due to a generalised inhibition of protein synthesis in SV3T3 cells was considered. Doses of interferon up to 500 units ml⁻¹ had no measurable effect on incorporation of ³H-amino acids in SV3T3 cells (Table 1a). The total amount of radioactivity (c.p.m.) incorporated was within the linear range for the experimental conditions used. Under similar conditions, cycloheximide (0.1–0.5 µg ml⁻¹) caused a marked, dose-dependent inhibition of ³H-amino acid incorporation (Table 1b). Both HF and cell lysate PA activities were also inhibited by cycloheximide. Thus, although inhibition of protein synthesis with cycloheximide causes decreased PA activity in SV3T3 cells, the interferon-mediated inhibition of PA release cannot be explained in terms of a general inhibition of protein synthesis.

Our data suggest that interferon inhibits some aspect of the process of release of PA. Although the mechanism of PA release is not clear, a secretion mechanism has been suggested^{17,18}. The firm membrane association of PA (refs 19–21), and its enrichment in plasma membrane fractions of chicken and hamster cells²¹, suggest to us that PA may be released through the process of membrane shedding, such as that described for histocompatibility antigens^{22,23} and certain membrane-associated glycosyl transferases^{24,25}. Thus, inhibition of PA release could be a direct effect of interferon on shedding. Alternatively, since PA release is dependent on cell density in 3T3 cells¹⁵, the interferon effect on SV3T3 cells may be indirect, because of inhibition of cell proliferation. Neither do we yet know whether the interferon effect extends to the release of other cell components. Experiments are in progress to examine these questions.

We thank Deborah DiGregorio for technical assistance. E.W.S. and S.J. are recipients of National Research Service Awards (1 F32 CA05975 and 1 F32 CA05913) from the National Cancer Institute. I.N.C. holds a Cancer Research Scholar Award from the American Cancer Society, Massachusetts Division. This research was supported in part by Grant CA10126 from the NCI, National Institutes of Health.

EDWARD W. SCHRODER
IHH-NAN CHOU
SUSAN JAKEN
PAUL H. BLACK

Department of Medicine,
Massachusetts General Hospital
and
Department of Medicine,
Harvard Medical School,
Boston, Massachusetts 02114

Received 31 July; accepted 24 October 1978.

- Gresser, I., Brouty-Boye, D., Thomas, M. & Macieira-Coelho, A. *Proc. natn. Acad. Sci. U.S.A.* **66**, 1052–1058 (1970).
- Kuwata, T., Fuse, A. & Morinaga, N. *J. gen. Virol.* **33**, 7–15 (1976).
- Strander, H. & Einhorn, S. *Int. J. Cancer* **19**, 468–473 (1977).
- Lindahl-Magnusson, P., Leary, P. & Gresser, I. *Proc. Soc. exp. Biol. Med.* **138**, 10450 (1971).

- Billiau, A., Hermans, H. & Allen, P. T. *Virology* **73**, 537–542 (1976).
- Chang, E. H., Myers, M. W., Wong, P. K. Y. & Friedman, R. M. *J. gen. Virol.* **34**, 363–367 (1977).
- Lindahl, P., Gresser, I., Leary, P. & Tovey, M. *Proc. natn. Acad. Sci. U.S.A.* **73**, 1284–1287 (1976).
- Lindahl, P., Leary, P. & Gresser, I. *Proc. natn. Acad. Sci. U.S.A.* **70**, 2785–2788 (1973).
- Chang, E. H., Jay, F. T. & Friedman, R. M. *Proc. natn. Acad. Sci. U.S.A.* **75**, 1859–1863 (1978).
- Hynes, R. O. & Bye, J. M. *Cell* **3**, 113–120 (1974).
- Hynes, R. O., Wyke, J. A., Bye, J. M., Humphries, K. C. & Pearlstein, E. S. in *Proteases and Biological Control* (eds Reich, E., Rifkin, D. B. & Shaw, E.) 931–944 (Cold Spring Harbor Laboratory, New York, 1976).
- Olden, K. & Yamada, K. M. *Cell* **11**, 957–969 (1977).
- Hirsch, M. S., Ellis, D. A., Black, P. H., Monaco, A. P. & Wood, M. L. *Transplantation* **17**, 234–236 (1974).
- Ogburn, C. A., Berg, K. & Paucker, K. J. *Immun.* **111**, 1206–1218 (1973).
- Chou, I. N., O'Donnell, S. P., Black, P. H. & Roblin, R. O. *J. cell. Physiol.* **91**, 31–37 (1977).
- Falcoff, R., Scheps, R., Sanceau, J., Catinot, L. & Falcoff, E. *Biochimie* **57**, 1109–1111 (1975).
- Unkeless, J. C., Tobia, A., Ossowski, L., Quigley, J. P., Rifkin, D. B. & Reich, E. *J. exp. Med.* **137**, 85–111 (1973).
- Ossowski, L., Unkeless, J. C., Tobia, A., Quigley, J. P., Rifkin, D. B. & Reich, E. *J. exp. Med.* **137**, 112–126 (1973).
- Unkeless, J., Dano, K., Kellerman, G. M. & Reich, E. *J. biol. Chem.* **249**, 4295–4305 (1974).
- Christman, J. K., Acs, G., Silagi, S. & Silverstein, S. C. in *Proteases and Biological Control* (eds Reich, E., Rifkin, D. B. & Shaw, E.) 827–839 (Cold Spring Harbor Laboratory, New York, 1975).
- Quigley, J. P. *J. Cell Biol.* **71**, 472–486 (1976).
- Kapeller, M., Gal-oz, R., Grover, N. B. & Doljanski, F. *Expl. Cell Res.* **79**, 152–158 (1973).
- Doljanski, F. & Kapeller, M. *J. theor. Biol.* **62**, 253–270 (1976).
- Bernacki, R. & Kim, U. *Science* **195**, 577–580 (1977).
- LaMont, J. T., Gammon, M. T. & Isselbacher, K. J. *Proc. natn. Acad. Sci. U.S.A.* **74**, 1086–1090 (1977).
- Finter, N. B. *J. gen. Virol.* **5**, 419–427 (1969).
- Chou, I., Roblin, R. O. & Black, P. H. *J. biol. Chem.* **252**, 6256–6259 (1977).
- Duetsch, D. G. & Mertz, E. T. *Science* **170**, 1095–1096 (1970).
- Lowry, O. H., Rosebrough, N. J., Farr, A. L. & Randall, R. J. *J. biol. Chem.* **193**, 265–275 (1951).

Different rules govern help for cytotoxic T cells and B cells

RECENT experiments with radiation chimaeras^{1,2} indicate that T cells differentiating in an H-2D compatible, but H-2I different, thymic environment are completely unresponsive to vaccinia virus. The explanation favoured² for this surprising finding is that participation of helper T cells³ (TH) is an absolute requirement for the generation of virus-immune cytotoxic T cells (Tc). No response is seen because the TH populations are 'locked in' to recognising the spectrum of H-2I antigens presented on radiation-resistant thymic cells². They cannot, therefore, interact with the Tc precursors which are genetically different in the I region. The result is a total failure of responsiveness. We report here that we have tested this concept further by stimulating acutely tolerised^{4,5} T-cell populations in H-2I different, but H-2D compatible virus-infected recipients. Our results indicate that there is apparently no need for participation of TH populations restricted to the H-2I region specificities expressed in the recipient environment. This differs from the situation for TH-B-cell collaboration^{4,5}, and suggests that help for the Tc response operates directly between T-cell subsets.

An elegant analysis of genetic requirements for T-cell-B-cell collaboration has been made by Sprent^{4,5}, who used negative selection protocols to remove alloreactive lymphocytes and then tested for TH to sheep red blood cells in irradiated recipients compatible for different regions of the H-2 gene complex. The

Table 1 H-2 haplotypes of mice and target cells

	K	A	B	J	E	C	S	D
B10.Br, C3H, CBA/J, L cell	k	k	k	k	k	k	k	k
BALB/c, B10.D2, P815	d	d	d	d	d	d	d	d
C57BL/6, BALB/B, MC57G	b	b	b	b	b	b	b	b
B10.A	k	k	k	k	k	d	d	d
B10.A(2R), 2RSV	k	k	k	k	k	d	d	d
B10.A(3R)	b	b	b	b	k	d	d	d
B10.A(5R), 5RSV	b	b	b	k	k	d	d	d
BALB/G, B10.HTGSV	d	d	d	d	d	d	d	b

SV, Target cells transformed with SV40 virus.

conclusion reached from these studies, and from experiments with chimaeras, is that the macrophages in the stimulator environment and the B cells to be helped must share I-region determinants. Furthermore, as for the Tc response to vaccinia virus², differentiating TH precursors only operate in the context of H-2I antigens encountered previously in radiation-resistant thymus⁵.

Our intention was to determine whether the same spectrum of interactions applies to the generation of virus-immune cytotoxic T cells. We know that a detectable Tc response only occurs following exposure to virally modified macrophages expressing the H-2K or H-2D phenotype encountered previously in thymus. Does the postulated interaction of TH and Tc² cells also require involvement of an infected macrophage presenting, as for T-B collaboration^{4,5}, the same I-region determinants expressed on radiation-resistant thymus epithelium?

Similar negative selection procedures^{4,5} and H-2 haplotype combinations were thus used to analyse the postulated requirement for TH in the Tc response to vaccinia virus. B10.D2 spleen and lymph node populations were first filtered through an irradiated (BALB/c × C57)F₁ environment, and thoracic duct lymphocytes were then stimulated with vaccinia virus⁶ in B10.A(5R) recipients, which differ from H-2K to I-E. Contrary to expectations, a strong virus-specific Tc response was seen at H-2D^d (Table 2). This cannot be attributed to operation of radiation-resistant TH in the recipients, as the chimaera cells were exposed to virus in the same way but were unresponsive². Stimulation in the absence of other than H-2D end (I-C to -D) homology with the recipient was confirmed in a further experiment (Table 3), where (C3H × BALB/c)F₁ T cells were used to minimise the possible involvement of some form of allogeneic effect⁷ operating at I-J or I-E (Tables 1 and 2). Furthermore, the

addition of lymphocytes capable of mediating allogeneic effects during the sensitisation process (unpublished data) has not, to date, caused any significant enhancement of cytotoxicity in situations where there is a low Tc response⁸ to vaccinia virus.

A comparable lack of requirement for I-region homology between T-cell and irradiated recipient was also found for another H-2 haplotype combination (Table 4). However, in one case (Expt 2, Table 4), there was evidence of either residual alloreactivity or generation of a virus-immune cytotoxic T-cell response operating across the H-2 barrier. The second of these possibilities has now been confirmed⁹.

Apparently, virus-specific cytotoxic T cells do not need to share I-region genes with the recipient in which they are stimulated, although the presence of the relevant H-2K- or H-2D-encoded antigens is essential (Tables 2 and 4). Thus, if TH cells are involved in the Tc response to vaccinia virus, the mechanism seems to be somewhat different from that defined for TH-B-cell-macrophage interactions, where I-region restriction seems to be absolute when analysed by similar experimental protocols^{4,5}. Perhaps helper T cells involved in the virus-immune cytotoxic response are cross-reactive for infected macrophages expressing I-A^d and I-A^b (Tables 2 and 3), or I-A^d and I-A^k (Table 4). The latter is, however, the same combination for which a total failure of the Tc response to vaccinia virus was found in the chimaera experiments². An alternative interpretation is that the postulated interaction between TH and Tc is quite independent of the recipient macrophages and operates essentially between T-cell subsets. Recognition may, for example, be specific for T-cell receptor idotype¹⁰. We must also consider the possibility that the influence of I-region genes on the virus-immune Tc response is mediated other than through T-cell help at the time of antigen stimulation, perhaps reflecting

Table 2 Negatively selected T cells of the I-A^d genotype generate a strong cytotoxic response when stimulated in an H-2D^d-compatible recipient expressing I-A^b

Population stimulated	950R recipient	% specific ⁵¹ Cr release (25:1)					
		5RSV		P815		MC57G	
		Vacc.	N	Vacc.	N	Vacc.	N
B10.D2 in C57	B10.A(5R)	69	8	28	—	0	0
(BALB/c × C57)F ₁	B10.A(5R)	85	8	33	4	28	0
Unirradiated controls	(BALB/c × C57)F ₁	57	11	30	4	23	13
	CBA/J	26	23	0	0	0	14

B10.D2 spleen and lymph node cells were inoculated intravenously ($2.0\text{--}4.0 \times 10^8$ cells per mouse) into 10 irradiated (950R, 24 h) (BALB/c × C57BL/6)F₁ recipients, and negatively selected T-cells were drained from thoracic duct over the following 18–42 h (ref. 4). Approximately 4.0×10^7 lymphocytes were recovered from the 10 filter mice. These cells were then transferred to two irradiated recipients. Controls were unirradiated mice, or 950R recipients given 6.0×10^7 spleen and lymph node cells. All mice were injected intravenously with 5.0×10^6 plaque-forming units of vaccinia virus at 3 h after cell transfer, and spleen cells were tested for cytotoxicity 6 d later in a standard ⁵¹Cr release assay (ref. 11) using vaccinia-infected (vacc.) or normal (N) target cells. All values given are for specific ⁵¹Cr release, relative to incubation in medium alone. A strong cytotoxic T-cell response is seen at H-2D^d (5RSV) but not at H-2K^b (MC57G).

Table 3 Stimulation at H-2D^d in the absence of possible allogeneic effects

Population stimulated	950R recipient	% specific ⁵¹ Cr release (50:1)							
		5RSV		MC57G		P815		L cell	
		Vacc.	N	Vacc.	N	Vacc.	N	Vacc.	N
(C3H × BALB/c)F ₁ in	B10.A(5R)	61	12	10	6	23	1	—	—
(CBA × C57BL/6)F ₁	B10.A(3R)	59	0	11	0	21	0	2	—
Unirradiated controls	B10.A(5R)	71	10	57	0	26	1	16	116
	C57BL/6	67	11	64	2	9	6	25	23
	B10.A	49	11	5	3	19	1	81	11
	B10.D2	53	10	9	5	38	1	11	16
	B10.Br	11	7	5	0	7	4	77	8
	BALB/c	36	6	5	0	28	3	15	10
	(C3H × BALB/c)F ₁	38	12	17	9	19	1	55	12
	(BALB/c × C57BL/6)F ₁	56	6	31	0	28	1	13	13

(C3H × BALB/c)F₁ spleen and lymph node cells were filtered through 20 irradiated (CBA/J × C57BL/6)F₁ recipients and then stimulated with vaccinia virus in 950R recipients, as described in Table 2.

Table 4 Stimulation of negatively selected T cells in recipients which differ from K to I-J but share D-end antigenic specificities

Expt	Population stimulated	950R recipients	% specific ^{51}Cr release (25:1)			
			5RSV Vacc.	N	L cell Vacc.	N
1	BALB/c in CBA	A/J	32	-	1	4
	(BALB/c $\times$ CBA) F_1	BALB/c	68	11	7	0
	unirradiated CBA/J		6	8	75	0
2	BALB/c in CBA	B10.A	40	7	21	-
	(BALB/c $\times$ CBA) F_1	(BALB/c $\times$ CBA) F_1	61	4/	73	11
	(BALB/c $\times$ CBA) F_1	BALB/c	28	7	4	3
	(BALB/c $\times$ CBA) F_1	CBA	2	0	68	7

physiological mechanisms functioning during the process of lymphocyte differentiation in thymus^{1,11}. Is help involved in the generation of the Tc repertoire?

This work was supported by USPHS research grants CA-09140 and CA-20833 from the NCI and AI-14162 from the National Institute of Allergy and Infectious Diseases.

JACK R. BENNINK
PETER C. DOHERTY

The Wistar Institute of Anatomy and Biology,
36th Street at Spruce,
Philadelphia, Pennsylvania 19104

Received 26 April; accepted 31 October 1978.

1. Zinkernagel, R. M. *et al. J. exp. Med.* **147**, 882-896 (1978).
2. Zinkernagel, R. M. *et al. J. exp. Med.* **147**, 897-911 (1978).
3. Cantor, H. & Boyse, E. A. *J. exp. Med.* **141**, 1390-1399 (1975).
4. Sprent, J. & von Boehmer, H. *J. exp. Med.* **144**, 617-626 (1976).
5. Sprent, J. *Immun. Rev.* (in the press).
6. Bennink, J. & Doherty, P. C. *J. exp. Med.* **148**, 128-135 (1978).
7. Katz, D. H. & Benacerraf, B. *Adv. Immun.* **15**, 1-94 (1972).
8. Doherty, P. C., Biddison, W. E., Bennink, J. & Knowles, B. B. *J. exp. Med.* **148**, 534-543 (1978).
9. Doherty, P. C. & Bennink, J. R. *J. exp. Med.* (in the press).
10. Ward, K., Cantor, H. & Boyse, E. A. in *The Immune System: Genetics and Regulation* (eds. Sercarz, E., Herzenberg, L. A. & Fox, C. F.), 397 (Academic, New York, 1977).
11. Langman, R. E. *Nature* **274**, 12-13 (1978).

Suppression of human T-cell colony formation during pregnancy

THE human fetus, as a natural allograft, escapes destruction by the maternal immune system. As a possible mechanism, Alexander¹ proposed a rapid shedding of transplantation antigens by embryonic cells, resulting in a blindfolding of the otherwise fully functional maternal immune system. Other investigators have reported a suppressive activity of lymphoid cells from human cord blood on the proliferation of maternal lymphocytes^{2,3}. Suppressor T cells are induced by α -fetoprotein in mice⁴, and the elevated levels of α -fetoprotein during pregnancy have been shown to exert an immunosuppressive influence in mice^{5,6} and in man⁷. Olding⁸ demonstrated a soluble factor from mitogen stimulated lymphoid cells of human newborns suppressing the proliferation of maternal lymphocytes. These observations point out that the embryo might protect itself by release of soluble factors, acting either themselves as lymphocyte suppressive agents, and/or by the induction of suppressor cells. However, no direct evidence has so far been presented to show the presence and action of T-cell suppressor cells in the immune system of pregnant women. Such an experiment would have to use components of the immune system of pregnant women solely, and therein demonstrate an immune suppression as compared to non-pregnant individuals. We present here results indicating that the ability of T cells from pregnant women to proliferate in soft agar cultures is strongly inhibited, regardless of the source of the serum being used in the assay, thus pointing rather to a cell-mediated suppressor mechanism.

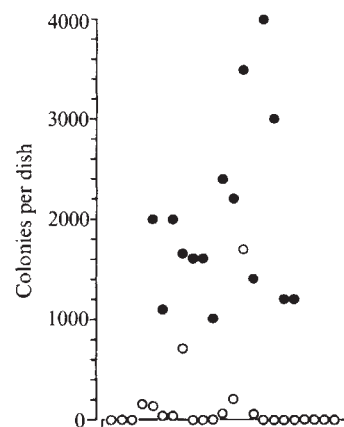


Fig. 1 3×10^6 lymphocytes were incubated overnight with PHA ($4 \mu\text{g ml}^{-1}$) in 5 ml RPMI 1640 containing 15% autologous serum. After 15-18 h, 2×10^5 cells were seeded into a RPMI-soft agar medium, containing the same serum, and incubated for 5 d at 37°C and 7% CO_2 . The number of T-cell colonies from non-pregnant donors (●) ranged from 1,000 to 4,000. The pregnant women (○) had a strongly impaired colony formation.

Twenty healthy pregnant women aged 20-34 yr and between the 10th and 41st weeks of gestation, gave their informed consent to *in vitro* analysis of cellular immune functions. Venous blood was collected in syringes containing RPMI 1640 and heparin, and the lymphoid cells were isolated by the Ficoll-Hypaque technique⁹.

Phytohaemagglutinin A (PHA) and concanavalin A (ConA) stimulation *in vitro* was performed according to Cunningham-Rundles¹⁰. T cell-quantitation¹¹ was performed in each case, as well as the T-cell colony test (TCC-test^{12,13}). For the TCC-test, purified PHA (Wellcome, catalogue no. HA 17) instead of PHA-M was used. Thus we could obtain 1,000-4,000 colonies per dish, seeding 2×10^5 lymphocytes into the dishes, from healthy non-pregnant donors. Of the cells obtained from the colonies, 80-90% formed E-rosettes, characteristic for T cells.

Table 1 shows the T-cell proportions and the lectin stimulation test in the pregnant women compared with a normal population. No significant change in the T-cell proportions and no marked inhibition of lectin stimulation was seen in the pregnant women. Figure 1 shows the TCC-test in the pregnant women, compared with the activity in 15 healthy non-pregnant donors. In 61% of the cases the pregnant women were completely unable to form colonies. In 30% of the cases, the colony forming activity was reduced to less than 5% of the activity in the control donors, the normal mean value being 2,000, (standard deviation, s_x , 900) per dish. If the cells from the pregnant women were incubated in a pooled human serum in which the control cells gave normal colony counts, no improvement of the colony forming activity could be observed (Fig. 2a).

Lymphocytes of healthy donors cultivated as an ABO compatible system in the presence of serum from the pregnant women (Fig. 2b) gave normal colony counts. Hence we assume that the factor depressing the clonal proliferation of T-cells in

Table 1 T-cell proportions and lectin stimulation test in pregnant and non-pregnant women

	Pregnancy		Control	
	$\bar{x}$	s_x	$\bar{x}$	s_x
T cells (%)	68.5	10.3	72.5	8.3
PHA stimulation (c.p.m.)	32,626	16,060	21,590	6,171
ConA stimulation (c.p.m.)	11,122	5,075	18,620	11,206

5×10^4 lymphocytes were incubated in 150 μl RPMI 1640 with 15% pooled human serum and $60 \mu\text{g ml}^{-1}$ ConA or $20 \mu\text{g ml}^{-1}$ PHA-P. After 96 h at 37°C , $0.2 \mu\text{Ci } ^3\text{H-TdR}$ (6.4 Ci mmol^{-1}) was added. After 24 h the cultures were collected using a multiple sample harvester, and counted in a scintillation counter.

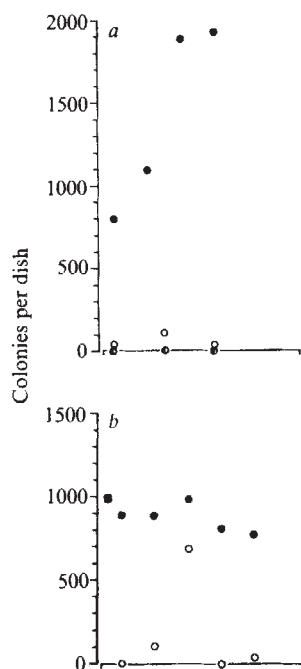


Fig. 2 *a*, T-cell colonies from non-pregnant donors in pooled serum (●), from pregnant women, incubated in the same serum pool (◐) and in autologous serum (○). Substitution of the serum from the pregnant women by normal serum caused no improvement of the colony formation. *b*, T-cell colony count from one non-pregnant cell donor after incubation in autologous serum (■), and in serum of ABO-compatible pregnant women (●). The T-cell colony formation from the same pregnant women in their own serum served as a control (○). No significant depression of normal T-cell colony formation was seen in serum from pregnant women.

this system is cellular, rather than humoral. The presence of a factor in the blood of pregnant women able to suppress the proliferation of T lymphocytes may, considering the report of Olding⁸, reflect the ability of the fetus to defend itself against the maternal immune system.

The failure of lectin stimulation analysis to show any significant change in the pregnant women indicates that the TCC-test has a much higher sensitivity in detecting factors inhibiting T-cell proliferation.

The kinetics of the suppressor phenomenon described here, are now being investigated. Preliminary results have shown that the immune depression described above is seen as early as 6–10 weeks after onset of pregnancy, and is abolished between the 35th week of gestation and the sixth week after delivery.

The work was supported by the Deutsche Forschungsgemeinschaft and the Deutsche Krebshilfe e.V.

ROLAND STAHN

Forschergruppe Tumorimmunologie,
Universität Freiburg, Stefan-Meier-Str 8,
D 7800 Freiburg, FRG

HANS-ÅKE FABRICIUS

Medizinische Universitätsklinik, Hugstetter Str 55,
D 7800 Freiburg, FRG

WOLFGANG HARTLEITNER

Universitäts-Frauenklinik, Hugstetter Str 55,
D 7800 Freiburg, FRG

Received 15 May; accepted 3 November 1978.

- Alexander, P. *Cancer Res.* **34**, 2077–2082 (1974).
- Olding, L. B. & Oldstone, M. B. A. *J. Immun.* **116**, 682–686 (1976).
- Oldstone, M. B. A., Tishon, A. & Moretta, L. *Nature* **269**, 333–335 (1977).
- Murgita, R. A., Goidl, E. A., Kontiainen, S. & Wigzell, H. *Nature* **267**, 257–259 (1977).
- Zimmermann, E. F., Voorting-Hawking, M. & Michael, J. G. *Nature* **265**, 354–356 (1977).

- Murgita, R. A. *Scand. J. Immun.* **5**, 1003–1014 (1976).
- Auer, I. O. & Kress, H. G. *Cell Immun.* **30**, 173–180 (1977).
- Olding, L. B., Murgita, R. A. & Wigzell, H. *J. Immun.* **119**, 1109–1114 (1977).
- Böyum, A. *Scand. J. clin. Lab. Invest.* **21**, Suppl. 97, 77 (1968).
- Cunningham-Rundles, S. *Clin. Immunobiology* Vol. 3 (eds Bach, F. H. & Good, R. A.) 151–195 (Academic, New York, 1968).
- Aiuti, F. *et al. J. clin. Immun. Immunopath.* **3**, 584–597 (1975).
- Rosenszajn, L. A., Shoham, D. & Kalechman, I. *Immunology* **29**, 1041–1055 (1975).
- Fibach, E., Gerassi, E. & Sachs, L. *Nature* **259**, 127–128 (1976).

Catecholamines of supposed inhibitory hypophysiotrophic function suppress action potentials in prolactin cells

THE anterior lobe of the pituitary gland has a uniquely important function in the endocrine system, for its six different endocrine cell types constitute the major endocrine link between the brain and the rest of the body. The brain is known to control this remarkable composite gland through specialised neurones lodged in the hypothalamus that elaborate and release hypophysiotrophic factors; these are carried by the hypophysial portal blood vessels to act on their respective target cells through stimulation or inhibition of secretory activity^{1,2}. However, the way in which these factors act at the cellular level is uncertain, although discharge of the various hormones of the anterior pituitary lobe is known to show calcium dependence and can be induced by depolarising concentrations of potassium^{3–5}; this pattern of endocrine behaviour was first recognised in adrenal chromaffin cells⁶ and is consistent with the calcium hypothesis of stimulus–secretion coupling⁷, which, in chromaffin and some other cells, is associated with electrical excitability^{8–13}. It is therefore of much interest that normal gland cells of the anterior pituitary lobe^{14,15}, like their neoplastic counterparts^{16–18}, have been found to generate action potentials with a calcium component indicative of influx of calcium ions; furthermore, the frequency of action potentials can be increased in some of these cells by thyrotropin releasing factor (TRF)¹⁴. This has prompted the suggestion that it may be through the initiation or modulation of action potentials that the brain, through the hypophysiotrophic factors, regulates secretion in the anterior pituitary¹⁴. However, the evidence which supports this scheme is scanty. The experiments with TRF were carried out on disaggregated anterior pituitary lobes of rats which contain six different endocrine cells, and the method did not allow identification of the cells stimulated by TRF; thus, it could only be supposed that they were the appropriate target cells¹⁴. Clearly, it is essential that evidence be obtained on the electrical behaviour of normal anterior lobe cells of defined type, and the effects thereon of specific and relevant hypophysiotrophic factors. Here we report that normal prolactin cells generate spontaneous action potentials which are suppressed by hypothalamic catecholamines known to inhibit prolactin secretion.

In mammals, the intermingling of the various endocrine cells in the anterior lobe of the pituitary impedes the correlation of electrical behaviour with cell type. However, in teleost fishes, the various anterior pituitary cells are segregated by functional type into distinct regions. In particular, the prolactin cells, which are most appropriate for our purpose, are grouped in a distinct lobe (rostral pars distalis) which is easy to identify and excise^{19–21}. Moreover, teleostean prolactin cells are a suitable model as they behave like their mammalian counterparts in important respects: their secretory activity *in vivo* is subject to inhibitory hypothalamic control apparently exerted by hypophysiotrophic catecholamines; and *in vitro*, when this control is absent, secretion increases greatly but can be restrained by catecholamines^{19–32}. It follows that if action potentials are intimately involved in regulating secretion, then prolactin cells should show spontaneous action potential activity *in vitro* and this should be inhibited by catecholamines. Our experiments confirm these expectations and demonstrate that the action

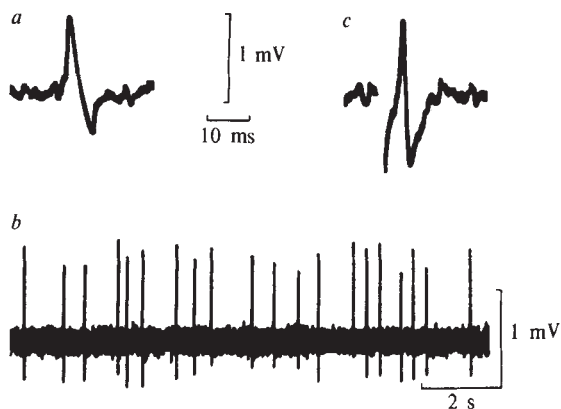


Fig. 1 Action potentials in prolactin cells. *a* and *b*, Spontaneous action potentials recorded extracellularly (with a microaspiration electrode¹⁴) at fast and slow sweep speeds from the same cell, to demonstrate the shape of individual action potentials (*a*) and pattern of spontaneous activity (*b*). *c*, Evoked action potential from a second cell elicited by an outward (depolarising) current pulse of 3.5 nA passed through the extracellular recording electrode. The frequency of spontaneous action potentials in this cell was low (0.08 Hz). The recording solution contained (in mM) NaCl, 183; KCl, 5; CaCl₂, 2; glucose, 5.5; HEPES, 5 (pH 7.1) with 1 mg ml⁻¹ bovine serum albumin (BSA). Experiments were done at 17–20 °C. Top calibrations apply to both *a* and *c*. The cells used for electrophysiological study were prolactin cells in primary culture obtained as follows. Pituitary glands were removed immediately after decapitating live, mature (about 30 cm) *Alosa pseudoharengus* (10–20 specimens per experiment). The anterior one-half to two-thirds of rostral pars distalis was excised from each, pooled and mechanically disaggregated by shearing with siliconised Pasteur pipettes with flame-polished tips of inner diameter ≤ 1 mm. Disaggregation medium was 3 ml Ham's F-10 (Gibco) supplemented with: 15% horse serum, 2.5% fetal calf serum, 35 mM NaCl, 100 U ml⁻¹ penicillin and 100 μ g ml⁻¹ streptomycin, and saturated with 5% CO₂–95% O₂ at room temperature. Dispersed cells were centrifuged at 100g for 8 min and the pellet resuspended in 2 ml recording solution (see above). To remove debris, suspended cells were layered on 5 ml recording solution containing 4% BSA and spun down at 100g for 8 min. Sedimented cells were resuspended in supplemented Ham's F-10 and placed in Falcon dishes (35 mm) coated with poly-L-lysine (Sigma type V). Dishes, each containing cells from two pituitaries in 2 ml medium, were maintained at room temperature and generally used for experiment within 10 d.

potentials involve a calcium component which probably provides a stimulus to secretion.

The fish used was the alewife (*Alosa pseudoharengus*); the pituitary gland of this species has a large and clearly defined rostral pars distalis shown to consist essentially of a mass of large prolactin cells containing prominent secretory granules, the only other endocrine cells present being small ACTH cells arranged in a thin layer along the posterior border³³. We avoided these ACTH cells by using only the anterior one-half to two-thirds of the rostral pars distalis, which we disaggregated and maintained in culture (see legend to Fig. 1). The dispersed cells were observed under phase contrast at high magnification ($\times 800$) and records were obtained from those whose large size and granular content indicated them to be prolactin cells.

Most of the cells monitored showed spontaneous action potential activity (Fig. 1*a*). Of >100 spontaneously active cells, almost all fired at random intervals (Fig. 1*b*), but a few fired in a bursting pattern. Action potentials could also be induced, or their frequency increased, by outward (depolarising) current pulses passed through the electrode (Fig. 1*c*). When driven by current pulses, cells sometimes fired at frequencies in excess of 25 Hz. The average rate of spontaneous discharge, however, was 1 Hz (0.8 ± 0.1 Hz, mean $\pm$ s.e.m., $n = 23$, records selected to avoid influences of drugs).

Tetrodotoxin (TTX) reversibly depressed the amplitude of the action potentials (Fig. 2*a, b*), indicating that they result

mainly from an increase in membrane conductance to sodium ions³⁴. However, a calcium component was also present, as blockade of potential-dependent potassium conductance by the addition of tetraethylammonium (TEA) to the TTX-containing solution resulted in large regenerative potentials that were reversibly suppressed by replacing calcium in the bathing solution with manganese (Fig. 2*c, d*). The action potentials thus involve inward currents of both sodium and calcium ions (see refs 35, 36).

The catecholamines dopamine and noradrenaline, which are the principal factors in hypothalamic extracts capable of inhibiting secretion from isolated prolactin cells^{37,39}, had inhibitory effects on spontaneously firing prolactin cells and either slowed or arrested the discharge of action potentials. Arrest was the most common response when these substances were used in a concentration of 10^{-6} M. It occurred within a few seconds and often persisted for many seconds after the catecholamine had been withdrawn (Fig. 3*a, b*). We have not attempted to determine minimal effective concentrations but have noted that concentrations as low as 10^{-8} M can slow impulse discharge (Fig. 3*c*). When discharge persisted, albeit at reduced frequency, during exposure to dopamine or noradrenaline, the amplitude of the action potentials was unaltered. The catecholamines therefore seem to inhibit the mechanism which initiates the action potentials (the 'generator potential') and not the potential-dependent conductances that underlie these spikes.

Our experiments thus provide new information on the control of anterior pituitary secretion. (1) They indicate that normal prolactin cells, in conditions in which they secrete spontaneously, have spontaneous action potentials; (2) they demonstrate that these action potentials, although mediated mainly by sodium, involve inward calcium movement which, from evidence on many secretory systems^{7,40} including prolactin and other anterior pituitary cells³⁻⁵, should provide a stimulus to secretion; (3) they show that putative inhibitory hypophysiotrophic factors reduce this electrical activity. Together, these findings strengthen the hypothesis that hypophysiotrophic factors control secretion in the anterior pituitary gland by

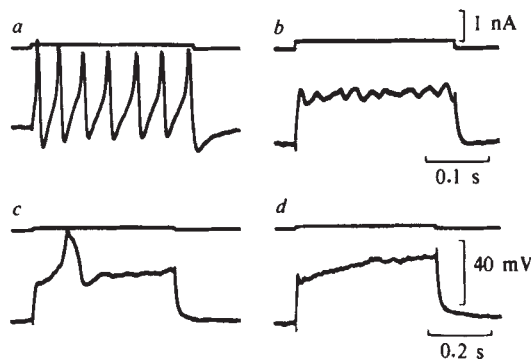


Fig. 2 Ionic basis of action potentials in prolactin cells. Intracellular recordings from two different cells showing the effect of TTX (*a, b*) and of replacement of Ca²⁺ by Mn²⁺ (*c, d*). *a*, Repetitive action potentials evoked by a single outward (depolarising) current pulse passed through the recording electrode. *b*, Response of the same cell during exposure to TTX (2×10^{-6} M) applied by a delivery pipette¹⁴. The action potentials are virtually abolished. *c*, Response of a second cell to a depolarising pulse while bathed in a solution containing TTX (5×10^{-6} M) with the further addition of TEA (10 mM). A prominent regenerative potential is evident. *d*, Response of the same cell when perfused (by drug delivery pipette) with the same medium except that Ca²⁺ was replaced by Mn²⁺. The regenerative response is abolished. Top trace in all records indicates current passed and zero level of membrane potential. Calibrations for current and potential apply to all four sets of records. Intracellular recording was done by conventional techniques using glass microelectrodes filled with 4M potassium acetate, pH 7.1, with resistance 60–100 M Ω . A bridge circuit was used for passing current. To facilitate intracellular recording the Ca²⁺ concentrations were increased to 10 mM throughout except in *d*, where it was replaced by 10 mM Mn²⁺.

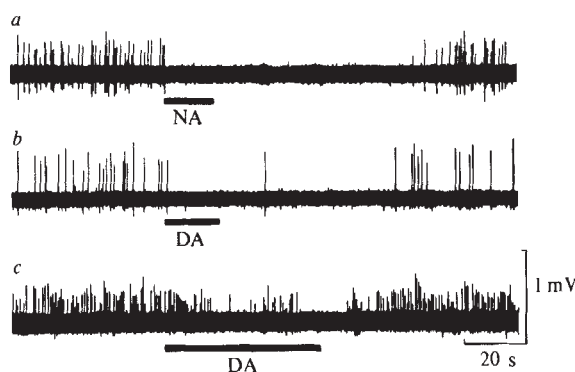


Fig. 3 Inhibition of spontaneous action potentials in prolactin cells by catecholamines. Catecholamines (in recording solution) were applied to the cells by drug delivery pipette for periods indicated by bars under the records. *a*, Arrest of spontaneous discharge by noradrenaline (10^{-6} M NA). *b*, Similar effect, in a second cell, produced by dopamine (10^{-6} M DA). *c*, Slowing of spontaneous discharge, in a third cell, by a lower concentration of dopamine (10^{-8} M). Similar responses were obtained in 32 of 34 cells tested.

affecting action potential activity in their target cells. Finally, in common with others¹⁹⁻²¹, we consider that the teleost pituitary, because of the segregation of cell types, offers unique advantages for the study of hypophysial function.

This work was supported by a grant from the USPHS.

P. S. TARASKEVICH
W. W. DOUGLAS

Department of Pharmacology,
Yale University School of Medicine,
New Haven, Connecticut 06510

Received 10 October; accepted 6 November 1978.

- Harris, G. W. *J. Endocr.* **53**, ii-xxiii (1972).
- Reichlin, S., Baldessarini, R. J. & Martin, J. B. (eds) *The Hypothalamus*, Vol. 56 (New York: Raven 1978).
- Vale, W., Rivier, G. & Brown, M. A. *Rev. Physiol.* **39**, 473-527 (1977).
- Wakabayashi, K., Date, Y. & Tamaoki, B. *Endocrinology* **92**, 698-704 (1973).
- Nakano, H., Fawcett, C. P. & McCann, S. M. *Endocrinology* **98**, 278-288 (1976).
- Douglas, W. W. & Rubin, R. P. *J. Physiol., Lond.* **159**, 40-57 (1961).
- Douglas, W. W. *Br. J. Pharmac.* **34**, 451-474 (1968); *Ciba-Symposium* **54**, 61-87 (1978).
- Biales, B., Dichter, M. & Tischler, A. *J. Physiol., Lond.* **262**, 743-753 (1976).
- Brandt, B. L., Hagiwara, S., Kidokoro, Y. & Miyazaki, S. *J. Physiol., Lond.* **263**, 417-439 (1976).
- Katz, B. *The Release of Neural Transmitter Substances* (Thomas, Springfield, 1969).
- Dean, P. M. & Matthews, E. K. *J. Physiol., Lond.* **210**, 255-264 (1970).
- Meissner, H. P. & Schmelz, H. *Pflügers Arch. ges. Physiol.* **351**, 195-206 (1974).
- Douglas, W. W. *Handbook of Physiology* Vol. 4 (eds Greep, R. O. & Astwood, E. B.) 191-224 (1974).
- Taraskevich, P. S. & Douglas, W. W. *Proc. natn. Acad. Sci. U.S.A.* **74**, 4064-4067 (1977).
- Ozawa, S. & Sand, O. *Acta physiol. scand.* **102**, 330-341 (1978).
- Kidokoro, Y. *Nature* **258**, 741-742 (1975).
- Biales, B., Dichter, M. A. & Tischler, A. *Nature* **267**, 172-174 (1977).
- Douglas, W. W. & Taraskevich, P. S. *J. Physiol., Lond.* **272**, 41-43P (1977).
- Sage, M. & Bern, H. A. *Int. Rev. Cytol.* **31**, 339-376 (1971).
- Schreibman, M. P., Leatherland, J. F. & McKeown, B. A. *Am. Zool.* **13**, 719-742, (1973).
- Holmes, R. L. & Ball, J. N. *The Pituitary Gland. A Comparative Account*, 170-220 (Cambridge University Press, 1974).
- McKeown, B. A. *Experientia* **28**, 675-676 (1972).
- Båge, G., Ekengren, B., Fernholm, B. & Fridberg, G. *Acta zool.* **55**, 25-45 (1974).
- Zambrano, D., Clarke, W. C., Hawkins, E. F., Sage, M. & Bern, H. A. *Neuroendocrinology* **13**, 284-298 (1974).
- Nagahama, Y., Nishioka, R. S., Bern, H. A. & Gunther, R. L. *Gen. comp. Endocr.* **25**, 166-188 (1975).
- Oliverau, M. *Gen. comp. Endocr.* **26**, 550-561 (1975).
- Peter, R. E., McKeown, B. A. *Gen. comp. Endocr.* **25**, 153-165 (1975).
- Zambrano, D. *Cell Tissue Res.* **61**, 551-563 (1975).
- Wigham, T., Ball, J. N. & Ingleton, P. M. *J. comp. Physiol. B* **104**, 87-96 (1975).
- Wigham, T., & Ball, J. N. *J. comp. Physiol. B* **110**, 135-143 (1976).
- Wigham, T., Nishioka, R. S. & Bern, H. A. *Gen. comp. Endocr.* **32**, 120-131 (1977).
- Oliverau, M. *Cell Tissue Res.* **189**, 219-230 (1978).
- Cook, H., Rusthoven, J. J. & Vogelzang, N. H. *Z. Zellforsch. mikrosk. Anat.* **414**, 145-159 (1973).
- Narahashi, T. *Physiol. Rev.* **54**, 812-889 (1974).
- Katz, B. & Miledi, R. *J. Physiol., Lond.* **203**, 459-487 (1969).
- Hagiwara, S. *Adv. Biophys.* **4**, 71-102 (1973).
- MacLeod, R. B. in *Frontiers in Neuroendocrinology* (eds Ganong, W. F. & Martini, L.) 169-194 (Raven, New York, 1976).
- Schally, A. V. *et al. Acta endocr., Copnh.* **82**, 1-14 (1976).
- Juorio, A. V. *J. Neurochem.* **20**, 641-645 (1973).
- Rubin, R. *Calcium and the Secretory Process* (Plenum, New York, 1974).

Neurone-specific enolase is a molecular marker for peripheral and central neuroendocrine cells

NEURONE-SPECIFIC ENOLASE (NSE) is the most acidic brain isoenzyme of the glycolytic enzyme enolase (EC4.2.1.11) and has been shown to be homologous to the 14-3-2 protein isolated from bovine brain by Moore¹⁻³. Whereas NSE is exclusively localised in neurones in mammalian nervous tissue^{4,5}, another brain enolase isoenzyme termed non-neuronal enolase (NNE) is localised in glial elements⁵. As NSE and NNE are structurally, functionally and immunologically distinct isoenzymes that represent separate gene products^{6,7}, they are useful markers for cell classes in the nervous system. Although NNE is probably identical to liver enolase, NSE has to date been considered to be localised in neurones. We now report that NSE is also present in peripheral and central neuroendocrine cells, also termed amine precursor uptake and decarboxylation (APUD) cells⁸⁻¹⁰. Immunocytochemistry using the unlabelled antibody enzyme method of Sternberger¹¹ demonstrates that APUD cells in both laboratory rat and rhesus monkey (*Macaca mulatta*) stain positively with NSE antiserum. Using a sensitive radioimmunoassay (RIA) for NSE, it is possible to confirm their localisation in adrenal gland and to support the finding in other APUD-cell-containing tissues in rat, monkey and man. The results in human tissues suggest that the immunocytochemical localisation in rat and monkey will be valid for man and prove useful in the study of human diseases involving the APUD cell class.

Antisera to purified neurone-specific enolase isolated from rat (NSE-R) and human (NSE-H) brain were prepared in New Zealand white rabbits as previously described¹². The sera were not reactive with NNE. Antiserum to NSE-H reacts identically with monkey and human brain extracts, indicating that this serum cross-reacts quantitatively with monkey NSE (NSE-M). Animals were anaesthetised and perfused through the aorta first with normal saline and then with 4% paraformaldehyde-1% glutaraldehyde-0.2% picric acid in 0.1 M sodium acetate buffer, pH 6.0, with 1.5% sucrose. Tissues were fixed for 2-6 h and washed overnight. Monkey tissue was cut into 10-25 μ m thick sections on a vibratome or cryostat; rat tissue was imbedded in polyester wax¹³ and cut at a thickness of 4-8 μ m. The rest of the staining proceeded according to the unlabelled antibody enzyme method¹¹. Preimmune serum and NSE-absorbed serum served as controls.

Stained sections revealed that NSE antiserum reacts specifically and uniformly with the APUD series of neuroendocrine cells⁸⁻¹⁰ in each gland or tissue (Table 1). Both rat and rhesus monkey gave identical results. In pancreas, the cells of the islets of Langerhans exhibit positive staining, in contrast to the surrounding exocrine acinar and epithelial cells, which do not stain (Fig. 1a). All islet cells stain positively, although there is some variation in intensity among cells. NSE staining in the adrenal gland (Fig. 1b) clearly delineates the positive chromaffin cells of the medulla from the non-staining cortex. NSE-positive cells in the thyroid (Fig. 1c) are compatible in position and number with parafollicular calcitonin (CT) cells, whereas the thyroid follicular cells are NSE negative. Likewise, pinealocytes in the pineal and cells of the pituitary, especially those of the pars intermedia, all stained for NSE. Interstitial and epithelial cells remained unstained. In the hypothalamus, neuroendocrine cells contain NSE as do all neurones in the central and peripheral nervous system.

To test the hypothesis that neuroendocrine cells contain NSE, quantitation of NSE in rat, monkey and human tissues was accomplished by RIA. Previous work has demonstrated that NSE is a major soluble protein constituent of brain¹². The levels of NSE in rat, monkey and human whole brain extracts range from 12,500 to 16,000 ng NSE per mg soluble protein (Table 1), representing 1.25-1.6% of the total soluble protein in brain.

Table 1 Neurone-specific enolase levels in various tissues of rat, rhesus monkey and man

Tissue	Rat	NSE (ng per mg protein) Monkey	Human	NSE Localisation by immunocytochemistry
Brain	12,500 ± 105	14,000	16,000	Neurones including central neuroendocrine cells of the hypothalamus
Pineal	2,650 ± 400	7,700	8,060	Pinealocytes
Adrenal	250 ± 11	—	—	—
Cortex	—	50	100	None
Medulla	—	680	940	Chromaffin cells
Pituitary	2,000	—	600	—
Anterior lobe	900	1,180	—	Anterior pituitary cells
Posterior lobe	3,400	2,870	—	Pars intermedia cells and endings of the hypothalamo-hypophyseal system
Pancreas	18 ± 1.5	49	—	Islet cells
Thyroid	150	—	250	Presumed CT cells
Muscle	9	<1	—	None
Liver	4	1	3	None

Rat tissues such as pineal, adrenal, pituitary and thyroid represent pools of 5–20 animals. All tissue was homogenised in 20 mM Tris-PO₄, 2 mM MgSO₄, pH 7.4. The 100,000g supernatant was obtained. A double-antibody radioimmunoassay was developed for NSE from rat, monkey and human. Antiserum raised in rabbits to purified NSE-H was used in the monkey and human tissue assays and antiserum to purified NSE-R for the rat tissue. Each assay contained 4,000 d.p.m. of tritiated antigen (20–30 ng), prepared as previously described¹², antiserum (1:40,000 dilution for anti-NSE-R and 1:1,000 for anti-NSE-H) and the sample to be measured. The assay volume was 0.5 ml and all solutions were diluted in the assay buffer (0.15 M Tris-PO₄, pH 7.4, 2 mM MgSO₄ and 1% bovine serum albumin). The samples were incubated at 4 °C for 16 h followed by the addition of 250 µl of a 2 mg ml⁻¹ solution of solid-phase anti-rabbit IgG (Bio-Rad Immunogarb). The second antibody incubation proceeded for 2 h at room temperature followed by washing and counting of the beads. Standard curves for each assay were linear over 2–100 ng. No cross-reactivity was observed with NNE (1 µg). S.e.m. are presented when four or more groups of animals were assayed.

Liver and muscle have very low levels of NSE that probably reflect the contribution of peripheral nerve processes present in these tissues. In contrast, significant amounts of NSE are present in the glandular tissues listed in Table 1, including levels in the pineal that approach those of brain. In contrast to pineal and pituitary, pancreas and thyroid have fewer APUD cells but still have levels of NSE that are clearly higher than those of liver and muscle. Assay of NSE in the adrenal medulla and cortex directly confirms the immunocytochemical localisation (Table 1). NSE levels by RIA are 10-fold higher in the medulla than in the cortex, consistent with the positive NSE staining of adrenal medullary chromaffin cells and non-staining of cortical cells.

The striking and consistent staining of APUD cells in rat and monkey by NSE antiserum coupled with quantitative evidence that in rat, monkey and man this represents actual NSE content makes NSE a unique, generalised molecular marker for these

neurone-like cells. Whereas previous classification of APUD cells has had to rely on staining and physiological properties which occasionally proved difficult to document for a given cell class^{8–10}, the presence of NSE provides a simple and useful method for identifying members of this group. In cultured neuroblastoma cells (transformed cells of the APUD cell series), the levels of NSE correlate with the degree of differentiation induced by pharmacological manipulation¹⁴. NSE may, therefore, prove useful in the diagnosis and staging of tumours of APUD cells (apudomas) in humans.

The presence of a neuronal antigen in neuroendocrine cells further contributes to studies of the neurone-like character of these cells¹⁰. Not only do cells of the APUD classification have secretory properties similar to those of a neurone and, in most cases, a known associated peptide, they also exhibit excitable membrane properties in tissue culture¹⁵. As NSE has distinct

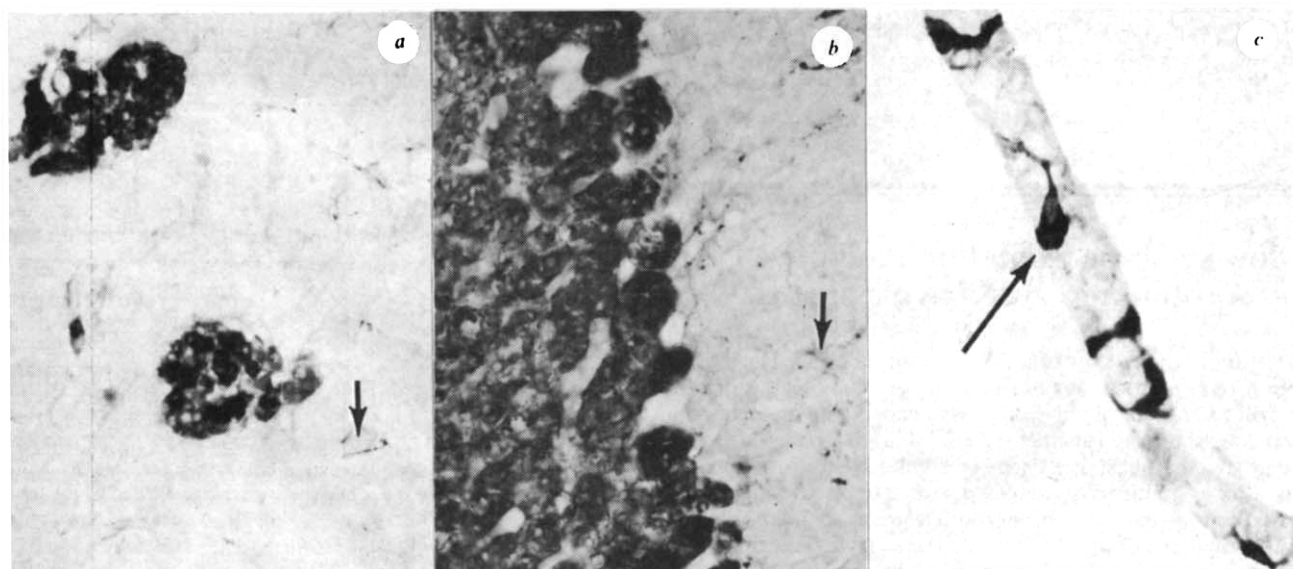


Fig. 1 Photomicrographs of rhesus monkey pancreas, adrenal and thyroid stained for NSE by the unlabelled antibody enzyme method. In each case, sections were incubated for 24 h with 1:1,000 anti-NSE antiserum. Controls carried out with preimmune and preabsorbed serum at 1:1,000 dilution showed no staining. *a*, 15-µm cryostat section of pancreas showing NSE staining of cells in islets of Langerhans. Some nerve processes are visible in the surrounding exocrine tissue (vertical arrow) but no epithelial or exocrine cells stain. ×130. *b*, 20-µm vibratome section of adrenal at the cortico-medullary border. Medullary chromaffin cells are stained and only occasional fragments of nerve processes stain in the cortex, but no cortical cells (vertical arrow). ×130. *c*, 15-µm cryostat section of thyroid showing the wall of one follicle with several NSE-positive cells (one is indicated by oblique arrow) amidst non-staining follicular cells. These positive cells are presumed CT cells. ×860.

structural and functional properties that may be an adaptation to the neuronal cytoplasm⁷, the presence of NSE may be related to the neurone-like function of APUD cells. Alternatively, NSE may reflect the neural crest and neuroectodermal origin of these cells during development. Whatever the reasons for its presence, the availability of a generalised molecular marker for both peripheral and central neuroendocrine cells will be useful in studies of cellular differentiation both *in vivo* and in tissue culture. Finally, the possibility of simplifying and enlarging the APUD cell classification may settle some of the controversy concerning peptide-secreting tumours and their classification in man.

We thank Drs Frederick K. Goodwin, Irwin J. Kopin and Ludwig Sternberger for their support during this work.

DONALD SCHMECHEL

Laboratory of Clinical Science,
National Institute of Mental Health

PAUL J. MARANGOS

Clinical Psychobiology Branch,
National Institute of Mental Health,
Bethesda, Maryland 20014

MILTON BRIGHTMAN

Laboratory of Neuropathology and Neuroanatomical Science,
National Institute of Neurological and Communicative Disorders
and Stroke,
Bethesda, Maryland 20014

Received 5 July; accepted 31 October 1978.

- Marangos, P. J., Zomzely-Neurath, C. & Goodwin, F. K. *J. Neurochem.* **28**, 1097–1107 (1977).
- Marangos, P. J. & Zomzely-Neurath, C. *Biochem. biophys. Res. Commun.* **68**, 1309–1316 (1976).
- Moore, B. W. in *Proteins of the Nervous System* (ed. Schneider, D. J.) (Raven, New York, 1973).
- Pickel, V. M., Reis, D. J., Marangos, P. J. & Zomzely-Neurath, C. *Brain Res.* **105**, 184–187 (1974).
- Schmechel, D., Marangos, P. J., Zis, A. P., Brightman, M. & Goodwin, F. K. *Science* **100**, 313–315 (1978).
- Marangos, P. J., Zis, A. P., Clark, R. L. & Goodwin, F. K. *Brain Res.* **150**, 117–133 (1978).
- Marangos, P. J., Parma, A. M. & Goodwin, F. K. *J. Neurochem.* **31**, 727–732 (1978).
- Pearse, A. G. E. *Proc. R. Soc. Biol. B.* **170**, 71–80 (1968).
- Pearse, A. G. E. *Z. Krebsforsch.* **84**, 1–18 (1974).
- Pearse, A. G. E. *Chromaffin, Enterochromaffin and Related Cells*, 147 (Elsevier, Amsterdam, 1976).
- Sternberger, L. A. *Immunocytochemistry* (Prentice-Hall, New Jersey, 1974).
- Marangos, P. J., Zomzely-Neurath, C. & York, C. *Archs Biochem. Biophys.* **170**, 289–293 (1975).
- Sidman, R. L., Mottla, P. A. & Feder, N. *Stain Technol.* **36**, 279–284 (1961).
- Marangos, P. J., Goodwin, F. K., Parma, A. M., Lauter, C. & Trams, E. *Brain Res.* **145**, 49–58 (1978).
- Tischler, A. S., Dichter, M. A., Biales, B. & Greene, L. A. *New Engl. J. Med.* **296**, 919–925 (1977).

Slow synaptic excitation mediated by serotonin in Auerbach's plexus

THE application of low concentrations of substance P to guinea-pig myenteric neurones has been reported by Katayama and North to produce depolarisation associated with an increase in input resistance and augmented excitability¹. On this basis, they suggested that substance P may be involved in the production of the slow excitatory postsynaptic potential (e.p.s.p.) which can be evoked in myenteric neurones by electrical stimulation of the interganglionic connectives². In this report, we show that application of serotonin (5-HT) mimics the slow e.p.s.p., and present several lines of evidence which suggest that 5-HT is a more likely candidate for the neurotransmitter of the slow e.p.s.p.

We used conventional methods to record intracellular electrical activity in myenteric ganglion cells of guinea-pig small intestine and to apply 5-HT iontophoretically to these cells². Thirty-one ganglion cells were studied during electrode impalements which were maintained for 2–8 h. The effects of

each drug and altered ionic composition of the bathing medium were always reversed by washing in normal Krebs solution, and all treatments were retested from two to four times on each cell.

The somal membranes of the 31 cells were relatively inexcitable in the absence of fibre tract stimulation and exogenous 5-HT. The cells would either not discharge in response to depolarising current pulses injected through the recording microelectrode or would fire only one, sometimes two, spikes only at the onset of the current pulse. When these spikes occurred, they were followed by prolonged (2–15 s) hyperpolarising after-potentials during which input resistance of the cells was decreased. These cells had resting potentials of –58 to –60 mV. Electrical stimulation of fibre tracts which entered the ganglia increased the excitability of the somal membranes of these cells. Increased excitability was reflected by maintained discharge of spikes that continued for 5–42 s after cessation of fibre tract stimulation (Fig. 1a), and by an increase in the number of spikes elicited by intracellular injection of a depolarising current pulse (Fig. 1b). In contrast to the relative inexcitability before fibre tract stimulation, immediately after and during such stimulation, spike discharge occurred throughout the application of depolarising pulses, and the frequency of spike discharge increased in direct proportion to the intensity of the injected current.

The increased excitability following fibre tract stimulation was associated with depolarisation of the cell membrane (–8 to –20 mV) and an increase in input resistance which persisted for 1–2 min after the end of the stimulation (Fig. 1c). During the excited state, the somal membranes discharged spikes at the offset of hyperpolarising current pulses, and these spikes as well as the continuous discharge were eliminated by hyperpolarisation of the membrane with steady hyperpolarising current injected through the recording micropipette (Fig. 1c). The increase in input resistance was not due to rectification because it still occurred when the depolarisation of the response was reversed by injection of hyperpolarising current.

The following observations indicated that the excitatory response to fibre tract stimulation resulted from release of a chemical substance. The response was (1) blocked in Krebs solution containing 16 mM Mg²⁺ and 1.25 mM Ca²⁺; (2) blocked in HEPES buffered Krebs solution containing 1 mM

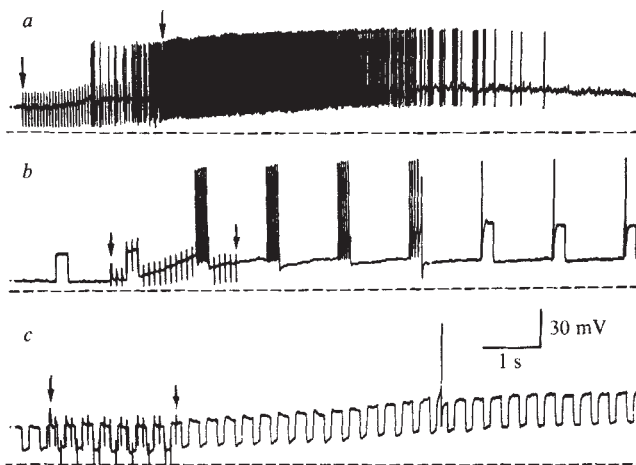


Fig. 1 Effects of fibre tract stimulation on guinea-pig myenteric neurones. *a*, Slow e.p.s.p. and spike discharge elicited by electrical stimulation of the interganglionic fibre tract (arrows indicate onset and offset of stimulus pulse train). *b*, electrical stimulation of fibre tract caused an increase in excitability as indicated by spike discharge throughout constant current depolarising pulses (0.4 nA). *c*, Fibre tract stimulation during repetitive injection of constant current hyperpolarising pulses (0.3 nA). Increased input resistance was reflected by increased amplitude of electrotonic potentials following fibre tract stimulation. Membrane potential was held 18 mV greater than resting potential by steady hyperpolarising current to prevent spikes at the offset of current pulses. One such spike still occurred.

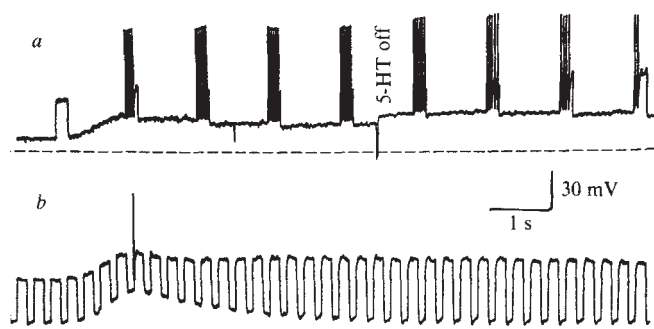


Fig. 2 Effects of microiontophoresis of 5-HT onto a guinea-pig myenteric neurone. *a*, Iontophoresis of 5-HT increased excitability and the cell discharged spikes throughout constant current depolarising pulses (4.5 nA). Iontophoretic current (300 nA) was turned on 9.5 s before the start of the record. Offset of the iontophoretic current is indicated. *b*, Iontophoresis of 5-HT during repetitive injection of constant current hyperpolarising pulses (4.5 nA). Increased input resistance was reflected by increased amplitude of electrotonic potentials during 5-HT iontophoresis. The membrane potential was held 20 mV greater than the resting potential by steady hyperpolarising current to prevent spikes at the offset of current pulses. One such spike still occurred. Iontophoretic current was turned on 4.5 s before the start of the record and remained on throughout. Separation between the tip of iontophoretic pipette and the recording electrode was $\sim 10 \mu\text{m}$. The tip of the iontophoretic pipette did not penetrate the periganglionic sheath.

Mn^{2+} ; (3) enhanced in Krebs solution containing 7.5 mM Ca^{2+} . The response was unaffected by 6 μM nicotine, 10 μM prostigmine, 140 μM (+)tubocurarine and 140 μM atropine which suggested that the chemical substance was not acetylcholine.

Microiontophoretic application of 5-HT mimicked the excitatory response to fibre tract stimulation (Fig. 2*a, b*). When 5-HT was iontophoreted onto the cells, the membrane potential depolarised (-6 to -20 mV), input resistance increased, the cells discharged spikes throughout a depolarising current pulse, and frequency of spike discharge during a current pulse was directly related to the intensity of the depolarising current. The alterations of membrane potential and input resistance required 1–3 min to return to control levels after removal of the iontophoretic current. Iontophoretic current passed through control pipettes containing 3 M NaCl had no effect on the neurones. The 5-HT solution in the iontophoretic pipettes had a pH of 3–4; however, it is unlikely that the action of 5-HT was due to pH change because the 5-HT was ejected into a relatively large volume buffered at pH 7.4.

The excitatory responses to both fibre tract stimulation and exogenous 5-HT were reversibly blocked by 30 μM methysergide, a 5-HT antagonist. The responses to both nerve stimulation and iontophoreted 5-HT were reversibly blocked after addition of 5-HT (1–2.5 μM) to the perfusion solution. The excitatory response to iontophoreted 5-HT was unaffected by the presence of 16 mM Mg^{2+} and 1 mM Mn^{2+} in the perfusion solution.

The slow synaptic potential in myenteric neurones is associated with reduction of postspike hyperpolarising potentials and a dramatic increase in the excitability of the somal membrane. Earlier workers referred to this population of myenteric neurones, which generates prolonged hyperpolarising after-potentials, as either type II cells³ or AH cells⁴, and suggested that they were sensory neurones, because no synaptic input could be demonstrated. This seems unlikely in view of the present findings.

The following observations by others have implicated 5-HT as a neurotransmitter within the myenteric plexus. (1) Some of the 5-HT effects on contractile activity of intestinal musculature *in vitro* do not occur after neural blockade with tetrodotoxin^{5,6}. (2) Extracellular recording from myenteric neurones shows an increase in the rate of spike discharge in the presence of 5-HT⁷.

(3) Application of 5-HT increases the rate of release of acetylcholine from myenteric neurones⁸. (4) Serotonin, tryptophan hydroxylase and serotonin binding protein are present within myenteric neurones^{9–12}. (5) Myenteric neurones synthesise 5-HT from its precursor, tryptophan¹². (6) Tetrodotoxin-sensitive release of tritium-labelled 5-HT and its binding protein occur during transmural electrical stimulation of intestinal segments *in vitro*¹³. (7) Myenteric neurones possess a high-affinity uptake mechanism for 5-HT^{14,15}.

When the present results are considered with previous evidence, it becomes apparent that most of the criteria for establishing 5-HT as a neurotransmitter in Auerbach's plexus have been fulfilled: serotonin is synthesised and released by myenteric neurones; microiontophoresis of 5-HT mimics the action of the natural transmitter; actions of both the natural transmitter and 5-HT are blocked by the 5-HT antagonist, methysergide; tachyphylaxis occurs in excess 5-HT; inactivation mechanisms in the form of specific uptake of 5-HT exist in the myenteric plexus.

Our observations together with those of Katayama and North¹ suggest that a functional relationship between substance P and 5-HT may exist, because substance P is also present within the myenteric plexus. It is noteworthy in this respect that in the brain 5-HT and substance P co-exist within synaptic vesicles¹⁶.

J. D. WOOD

Department of Physiology,
University of Kansas Medical Center,
Kansas City, Kansas 66103

C. J. MAYER

Physiologisches Institut
der Universität München,
Pettenkoferstr. 12,
D-8000 München 2, FRG

Received 6 June; accepted 10 November 1978.

1. Katayama, Y. & North, R. S. *Nature* **274**, 387–388 (1978).
2. Wood, J. D. & Mayer, C. J. *Pflügers Arch. ges. Physiol.* **374**, 265–275 (1978).
3. Nishi, S. & North, R. A. *J. Physiol., Lond.* **231**, 471–491 (1973).
4. Hirst, G. D. S., Holman, M. E. & Spence, I. *J. Physiol., Lond.* **236**, 303–326 (1974).
5. Bülbirg, E. & Gershon, M. D. *J. Physiol., Lond.* **192**, 823–846 (1967).
6. Burks, T. J. *Pharmac. exp. Ther.* **185**, 530–539 (1973).
7. Dingeldine, R., Goldstein, A. & Kendig, V. *Life Sci.* **14**, 2299–2309 (1974).
8. Vizi, V. A. & Vizi, E. S. *J. Neural Transmission* **42**, 127–138 (1978).
9. Fehér, E. & Csángi, K. *Acta anat.* **100**, 61–67 (1978).
10. Tafuri, W. L. & Raick, A. Z. *Natur.* **19**, 1126–1128 (1964).
11. Gershon, M. D., Dreyfus, C. F., Pickel, V. M., John, T. H. & Reis, D. J. *Proc. natn. Acad. Sci. U.S.A.* **74**, 3086–3089 (1977).
12. Jonakait, C. M., Tamir, H., Rapport, M. M. & Gershon, M. D. *J. Neurochem.* **28**, 277–284 (1977).
13. Jonakait, G. S., Tamir, H., Gintzler, A. R. & Gershon, M. D. *7th Int. Cong. Pharmac., Paris*, 737 (1978).
14. Dreyfus, C. F., Sherman, D. & Gershon, M. D. *Brain Res.* **128**, 109–123 (1977).
15. Gershon, M. D., Robinson, R. G. & Ross, L. L. *J. Pharmac. exp. Ther.* **198**, 548–561 (1976).
16. Chan-palay, V., Jonsson, G. & Palay, S. L. *Proc. natn. Acad. Sci. U.S.A.* **75**, 1582–1586 (1978).

Neurotransmitters decrease the calcium component of sensory neurone action potentials

RELEASE of neurotransmitters from presynaptic axon terminals requires the influx of Ca^{2+} ions during the nerve terminal action potential¹. Action potentials recorded in some neurone cell bodies exhibit a relatively large Ca^{2+} component, and it has been suggested that these soma Ca^{2+} spikes may provide a model for Ca^{2+} influx across the less accessible nerve terminal membrane². Recent data support the usefulness of this model. Serotonin (5-hydroxytryptamine, 5-HT) increases transmitter output at certain habituated sensory nerve-motoneurone synapses in the abdominal ganglion of *Aplysia* and it also prolongs the Ca^{2+} spike recorded in the sensory neurone cell body³. Enkephalin reduces the stimulated release

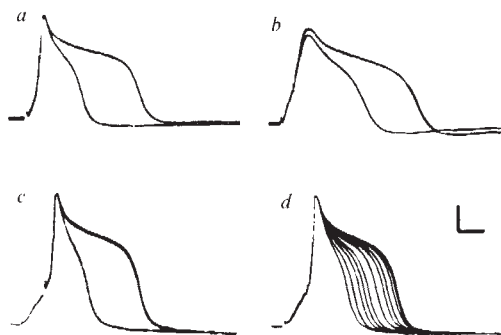


Fig. 1 The effect of *a*, 10^{-5} M 5-HT; *b*, 10^{-4} M GABA; *c*, 10^{-4} M NA on the DRG neurone soma action potential. The two successive sweeps show the spike before (longer duration) and after drug application. Return to control duration shown in *d* with 10-s interval between sweeps. Calibration, 20 mV and 2 ms.

of substance P by adult cat trigeminal neurones⁴ and by cultured embryonic chick dorsal root ganglion (DRG) neurones⁵, and it decreases the quantal content of excitatory postsynaptic potentials (e.p.s.ps, transmitter unknown) evoked in cultured rat spinal cord neurones by co-cultured DRG cells⁶. This peptide also decreases the duration and magnitude of the Ca^{2+} component of the DRG soma spike⁵. With the thought that modulation of Ca^{2+} currents may be a general correlate of presynaptic inhibition, we have studied the effect of several putative neurotransmitters on the soma spike of cultured chick sensory neurones, and report here that they decrease the calcium component of cell body action potentials.

Lumbar and thoracic DRG were dissected from 9–11-d-old chick embryos, incubated for 20 min in 0.01% collagenase, mechanically dissociated into single cells by trituration through a Pasteur pipette, and plated in 35-mm collagen-coated Falcon tissue culture dishes. Cells were maintained in Eagle's minimum essential medium supplemented with 10% horse serum, 5% chick embryo extract, 2 mM glutamine, 22 mM glucose, $5 \mu\text{g ml}^{-1}$ streptomycin, 50 IU ml^{-1} penicillin and $1.0 \mu\text{g ml}^{-1}$ of 2.5S nerve growth factor (NGF) (purified from mouse salivary gland). Non-neuronal background cell growth was halted at confluence by γ -irradiation with ^{60}Co .

Using previously described techniques⁷, intracellular recordings were made from DRG neurones which had been in culture for more than one week. The cells were bathed in a balanced salt solution containing 131 mM Na^+ , 5.9 mM K^+ , 5.4 mM Ca^{2+} , 0.8 mM Mg^{2+} , 0.1% glucose, 0.2% bovine serum albumin (BSA), and 25 mM HEPES, buffered to pH 7.4 with NaOH. The Ca^{2+} concentration was raised to 5.4 mM to increase inward Ca^{2+} currents and to improve the membrane seal around the microelectrodes. Known concentrations of drugs made up in recording medium were applied to individual neurones by pressure ejection from a 3–5- μm -tip micropipette positioned 100–200 μm from the impaled soma⁸.

The action potentials of embryonic chick DRG neurones are composed of a fast Na^+ component blocked by tetrodotoxin (TTX) and a slower Ca^{2+} component that is in part responsible for the plateau during repolarisation⁷. γ -Aminobutyric acid (GABA), noradrenaline (NA), and 5-HT produced a reversible dose-dependent shortening of this plateau phase of the action potential (Fig. 1). In one series, 97% ($n=36$) of the cells responded to GABA, 92% ($n=52$) to NA and 83% ($n=23$) to 5-HT. Dose-response curves for these drugs are shown in Fig. 2. The three drugs were found to exert their effects over approximately a 1,000-fold concentration range, but 5-HT and NA ($\text{ED}_{50}=6 \times 10^{-7}$ M) were slightly more potent than GABA ($\text{ED}_{50}=2 \times 10^{-6}$ M). With each drug, the maximum effect observed in control solution was a 60% reduction in spike duration measured at half amplitude.

Both the time course of drug action and the magnitude of the response were found to be dose related. Lower concentrations

($\sim 10^{-7}$ – 10^{-6} M) produced effects that lasted for several seconds, whereas return to control spike duration after application of higher drug concentrations (10^{-5} – 10^{-4} M) took several minutes.

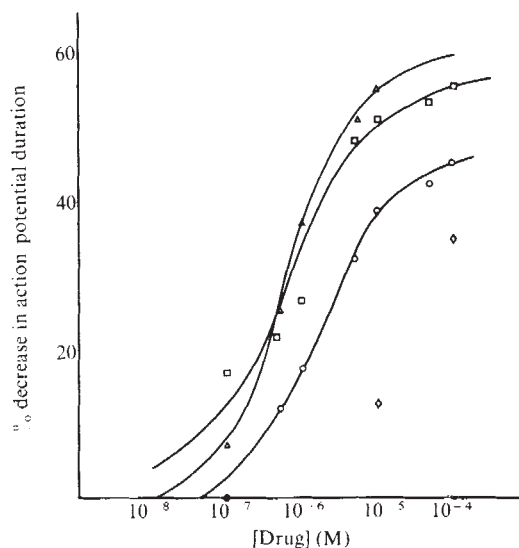
Glutamate produced a small and inconsistent effect on spike duration (Fig. 2). Only 8 of 17 cells responded to 10^{-5} or 10^{-4} M glutamate. ACh and glycine had no effect on the action potential in 10 neurones tested with 10^{-4} M. The fact that D-Ala-enkephalin amide and somatostatin at 10^{-6} M decreased DRG spike duration⁵ was confirmed. Other peptides including substance P, bradykinin, neurotensin and thyrotropin-releasing hormone were ineffective at 10^{-5} M.

Resting membrane potentials (V_m) of the DRG neurones ranged between -36 and -64 mV, with a mean of -50.7 ± 0.7 mV (s.e.m., $n=123$). Input resistance (R_{in}), measured as slope resistance through zero current, averaged ~ 50 M Ω . GABA, NA and 5-HT at concentrations of 10^{-7} – 10^{-4} M did not alter these resting membrane parameters. A small decrease in R_{in} following application of 10^{-4} M GABA to DRG neurones has been observed within the first week after plating⁸. In these young cultures, the change in conductance was small, less than 4% of that observed in dissociated spinal cord neurones. No change in R_{in} was produced by GABA in the older cultures (>1 week) used in this study. ACh at 10^{-4} M produced a 3–5 mV depolarisation associated with a small decrease in R_{in} .

We have begun to characterise the pharmacology of each response. The effect of 5×10^{-6} M GABA was not blocked by 10^{-4} M picrotoxin or 10^{-4} M bicuculline. These drugs are noncompetitive antagonists that block the GABA-mediated increase in Cl^- permeability at the crayfish neuromuscular junction⁹ and in the vertebrate nervous system^{10,11}. As no change in DRG resting membrane conductance was observed, it is not surprising that picrotoxin and bicuculline did not block the effect of GABA on the soma spike. Another kind of GABA receptor must be involved. The response to NA is apparently mediated by an α receptor: Selective α agonists, phenylephrine and dopamine at 10^{-5} M, were effective, whereas the β agonist, isoprenaline, at 10^{-5} M was not. Moreover, the effect of 10^{-5} M NA was blocked by 10^{-4} M phentolamine, an α antagonist, but not by 10^{-5} M propranolol, a β antagonist. Methysergide at 10^{-4} M did not block the effect of 10^{-5} M 5-HT.

The decrease in spike duration implies that the ultimate effect of GABA, NA and 5-HT is to decrease the duration of the inward Ca^{2+} current. It is not yet clear, however, if these drugs interact with Ca^{2+} channels directly. A decrease in spike duration might result from an increase in outward K^+ current.

Fig. 2 Dose-response curves for GABA ($\circ$), 5-HT (Δ) and NA ($\square$). Responses to high concentrations of glutamate are included ($\diamond$). Each point is the average per cent in spike duration of a minimum of five neurones.



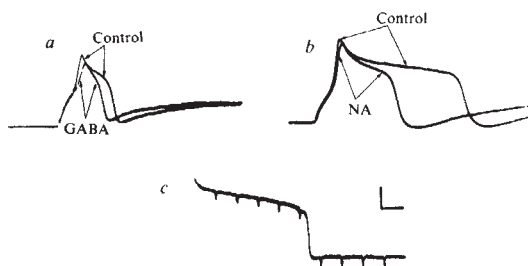


Fig. 3 *a*, Application of 10^{-5} M GABA decreases the rate of rise (dV/dt), overshoot (V_+) and duration of the Ca^{2+} spike recorded in 10^{-7} g ml $^{-1}$ TTX. *b*, Application of 10^{-4} M NA decreases dV/dt , V_+ and duration of the Ba^{2+} spike recorded in TTX and 3 mM Ba^{2+} . *c*, Conductance during the plateau of a Ba^{2+} spike (5 mM Ba^{2+}) measured with hyperpolarising current pulses shows only a small increase in plateau conductance over rest. Calibration, 20 mV and 5 ms (*a, b*) and 100 ms (*c*).

Voltage-dependent and Ca^{2+} -activated 12 K^+ currents must be considered. If this is the case, the outward current must occur early because GABA, NA and 5-HT also decrease the rate of rise and the peak amplitude of the Ca^{2+} spike (Fig. 3*a*). When the fast inward Na^+ current was blocked by 10^{-7} g ml $^{-1}$ TTX, GABA (10^{-5} M) decreased the maximum rate of spike rise from 28 ± 2.6 V s $^{-1}$ (mean $\pm$ s.e.m., $n = 10$) to 14 ± 2.5 V s $^{-1}$ ($n = 10$) and the spike overshoot from 11.5 ± 1.1 mV (mean $\pm$ s.e.m., $n = 10$) to 3.0 ± 1.8 mV ($n = 10$). Moreover, the same drug effects were observed in the presence of 3 mM Ba^{2+} (Fig. 3*b*). (Ca^{2+} was held constant at 5.4 mM.) Ba^{2+} passes through activated Ca^{2+} channels 13 and blocks outward K^+ currents 14 . DRG spikes were prolonged in Ba^{2+} and the membrane conductance tested during the long plateau was low (Fig. 3*c*), indicating that K^+ channels were, in fact, blocked. Thus, considering the effect of the drugs on spike rate of rise and the persistence of the effect in the presence of Ba^{2+} , we favour a direct drug effect on Ca^{2+} channels. More direct evidence must be obtained by varying extracellular ion concentrations. In any case, our findings indicate that neurotransmitters may modulate voltage-sensitive neuronal membrane channels without producing a change in resting membrane parameters. Neurotransmitter modulation of voltage-sensitive channels has been described in studies of cardiac $^{15-17}$ and smooth 18 muscle.

Frank and Fuortes were the first to describe a type of inhibition of 1a e.p.s.ps that was not associated with a detectable change in the postsynaptic (motoneurone) membrane conductance 19 . The weight of evidence suggests that this 'remote' inhibition is primarily, if not entirely, due to inhibition of transmitter release from afferent nerve terminals 20,21 but the precise mechanism of presynaptic inhibition remains obscure. Its temporal association with primary afferent depolarisation (PAD) has led to the notion that a transmitter-induced depolarisation of the primary afferent terminal reduces the size of the action potential and 1a transmitter output. GABA, the best studied neurotransmitter implicated in presynaptic inhibition, depolarises afferent nerve terminals (and nerve cell bodies) 11,22,23 by increasing the membrane permeability to Cl^- ions 11,23 . However, there is little direct evidence that this accounts for presynaptic inhibition.

Our results suggest an alternative mechanism. GABA decreases the influx of Ca^{2+} across the active soma membrane of cultured DRG neurones without producing a change in resting membrane conductance. If the same phenomenon occurs at nerve terminals this might effectively reduce transmitter release. It is interesting that this action of GABA is mimicked by 5-HT and NA because descending adrenergic (locus coeruleus) 24 and serotonergic (nucleus raphe magnus) 25 pathways have been described that terminate in the dorsal horn. Iontophoretically applied 5-HT and NA inhibit the firing of some neurones in superficial laminae of the dorsal horn 26 . Although the locus of this inhibition is not known, presynaptic actions of NA and 5-HT have been reported at other synapses 27,28 . It remains to be

seen which mechanism, decreased Ca^{2+} conductance or primary afferent depolarisation, is responsible for presynaptic inhibition *in vivo*.

We thank Deborah Youngman for the photography. Nerve growth factor was purified by Anne Mudge. This work was supported by USPHS grant NS11160 (G.D.F.) and USPHS training grant NS07009-03 (K.D.).

KATHLEEN DUNLAP
GERALD D. FISCHBACH

Department of Pharmacology,
Harvard Medical School,
250 Longwood Avenue,
Boston, Massachusetts 02115

Received 2 August; accepted 6 November 1978.

- Katz, B. & Miledi, R. *Proc. R. Soc. B* **161**, 496–503 (1965); **167**, 23–38 (1967); *J. Physiol., Lond.* **203**, 459–487 (1969).
- Stinnakre, J. & Tauc, L. *Nature new Biol.* **242**, 113–115 (1973).
- Klein, M. & Kandel, E. R. *Proc. natn. Acad. Sci. U.S.A.* **75**, 3512–3516 (1978).
- Jessell, T. M. & Iversen, L. L. *Nature* **268**, 549–551 (1977).
- Mudge, A., Leeman, S. & Fischbach, G. D. *Proc. natn. Acad. Sci. U.S.A.* (in the press).
- MacDonald, R. L. & Nelson, P. G. *Science* **199**, 1449–1451 (1978).
- Dichter, M. A. & Fischbach, G. D. *J. Physiol., Lond.* **267**, 281–298 (1977).
- Choi, D. thesis, Harvard Univ. (1978).
- Takeuchi, A. & Takeuchi, N. *J. Physiol., Lond.* **205**, 377–391 (1969).
- Johnston, G. A. R. in *Chemical Transmitters in the Mammalian Central Nervous System* (eds Hockman, E. & Bieger, D.) 31–83 (University Park Press, Baltimore, 1976).
- Gallagher, J. P., Higashi, H. & Nishi, S. *J. Physiol., Lond.* **275**, 263–282 (1969).
- Meech, R. & Standen, N. *J. Physiol., Lond.* **249**, 211–239 (1975).
- Hagiwara, S. *Adv. Biophys.* **4**, 71–102 (1973).
- Werman, P. & Grundfest, H. *J. gen. Physiol.* **44**, 997–1027 (1961).
- Vassort, G. *et al. Pflügers Archiv. ges. Physiol.* **225**, 263–267 (1969).
- Giles, W. & Noble, S. J. *J. Physiol., Lond.* **261**, 103–123 (1976).
- Tsien, R. W., Giles, W. & Greengard, P. *Nature new Biol.* **240**, 181–183 (1972).
- Bolton, T. B. *J. Physiol., Lond.* **250**, 175–202 (1975).
- Frank, K. & Fuortes, M. G. F. *Fedn Proc.* **16**, 39–40 (1957).
- Schmidt, R. F. *Ergebn. Physiol.* **63**, 20–101 (1971).
- Burke, R. E. & Rudomin, P. in *Handbook of Physiology: The Nervous System*, Vol. 1, 877–944 (Am. Physiol. Soc., 1977).
- Barker, J. L. & Nicoll, R. A. *J. Physiol., Lond.* **228**, 259–277 (1973).
- Nishi, S., Minota, S. & Karczmar, A. G. *Neuropharmacology* **13**, 215–219 (1974).
- Carlsson, A., Falck, B., Fuxe, K. & Hillarp, N. A. *Acta physiol. scand.* **60**, 112–119 (1964).
- Basbaum, A. I., Clanton, C. H. & Fields, H. L. *Proc. natn. Acad. Sci. U.S.A.* **73**, 4685–4688 (1976).
- Headley, P. M., Duggan, A. W. & Griensmith, B. T. *Brain Res.* **145**, 185–189 (1978).
- Breckenridge, B. McL., Burn, J. H. & Matschinsky, F. M. *Proc. natn. Acad. Sci. U.S.A.* **57**, 1893–1897 (1967).
- Dudel, J. *Arch. exp. Path. Pharmac.* **249**, 515–528 (1965).

Adenosine inhibits the accumulation of cyclic AMP in cultured brain cells

ADENOSINE stimulates the accumulation of cyclic AMP in some tissues and cell types, but has inhibitory or biphasic effects on the accumulation of cyclic AMP in other cells. Londos and Wolff 1 have attempted to explain these diverse effects by the existence of two adenosine reactive sites on adenylate cyclase. The first is the R-site, occupancy of which usually leads to activation on adenylate cyclase and which for activity requires integrity of the ribose ring. Much evidence suggests that this site is identical to an extracellular receptor for adenosine $^{2-6}$. The second type is the P-site, which mediates inhibition of adenylate cyclase and requires integrity of the purine ring for activity. Preliminary evidence suggests that this site is only accessible from the interior of the cell 1,3 . However, we report here that the inhibition by adenosine of the increase in the intracellular level of cyclic AMP evoked by isoprenaline in cultured glioblasts is mediated by extracellular receptors.

Brain cells from perinatal mice were obtained by mechanical dissociation of brain tissue and cultured as described previously 7 . The cultures consist mainly of epitheloid astroblasts 8 , which can be differentiated into cells exhibiting astrocyte-like appearance 9 . They are considered as models for glial cells 7,10,11 , as they express presumptive glial markers $^{12-15}$. The cultures were incubated for 10 min at 37 °C in the presence of the various additions before the concentration of intracellular cyclic AMP

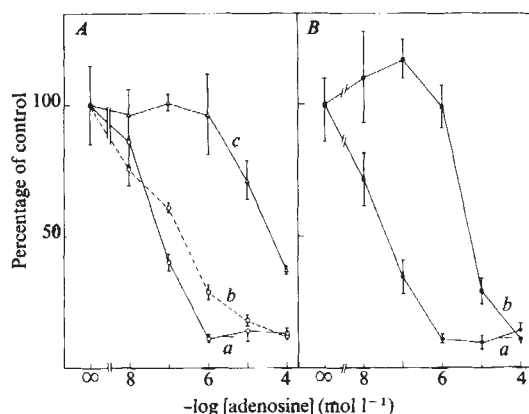


Fig. 1 Inhibition by adenosine of the elevation by isoprenaline (1 μ M) of the level of cyclic AMP in mouse brain cell cultures in the absence (A and B, curve a) or presence of 0.5 mM Ro20-1724 (A, curve b), 0.5 mM MIBX (A, curve c) or 0.5 mM theophylline (B, curve b). The culture for 24 d (A) or 20 d (B) of brain cells from perinatal mice (19–21 days *in utero*), the experimental incubation and the determination of cyclic AMP have been described elsewhere^{7,16}. Protein content²⁸: 1.0 mg per plate. Controls (pmol cyclic AMP per mg protein): A: isoprenaline (1 μ M), 580 ± 90 (= 100%); isoprenaline + Ro20-1724 (0.5 mM), $4,800 \pm 1,610$ (= 100%); isoprenaline + MIBX (0.5 mM), $4,700 \pm 290$ (= 100%); no addition, 7 ± 2 ; adenosine: 1 μ M, 8 ± 2 ; 10 μ M, 13 ± 4 ; 100 μ M, 16 ± 3 ; Ro20-1724 (0.5 mM), 24 ± 2 ; Ro20-1724 (0.5 mM) + adenosine: 1 μ M, 22 ± 4 ; 10 μ M, 35 ± 3 ; 100 μ M, 73 ± 13 ; MIBX (0.5 mM), 45 ± 4 (duplicates); MIBX (0.5 mM) + adenosine: 10 μ M, 44 ± 2 ; 100 μ M, 46 ± 6 . B: isoprenaline (1 μ M), 480 ± 70 (= 100%); isoprenaline + theophylline (0.5 mM), 830 ± 70 (= 100%). All data given are mean values of triplicates $\pm$ s.d., except where indicated.

was determined^{7,16}. Isoprenaline strongly elevates the level of cyclic AMP (ref. 7). This increase is inhibited by adenosine (Fig. 1A and B, curve a). Half maximal inhibition occurs at a concentration (IC_{50}) of 50–100 nM. The methylxanthines methylisobutylxanthine (MIBX) (Fig. 1A, curve c) and theophylline (Fig. 1B, curve b), known antagonists at adenosine receptors^{2,3,5,17–20} and inhibitors of phosphodiesterase (PDE)^{2,21}, prevent this inhibition. The PDE-inhibitors 4-(3-butoxy-4-methoxybenzyl)-2-imidazolidinone (Ro20-1724) (ref. 22) (Fig. 1A, curve b), dipyrindamole and papaverine^{2,21}, however, alone or in combination (not shown) do not substantially influence the inhibition by adenosine. The blockage by methylxanthines of the effect of adenosine is therefore due to their potency as antagonists at adenosine receptors and is not related to inhibition of PDE. Dipyrindamole is not only a PDE-inhibitor but also a strong blocker of adenosine uptake into cells^{2,4,23}. In the primary cultures investigated here, 100 μ M dipyrindamole inhibited the uptake of adenosine (concentration range 0.1–10 μ M) by 80 to 90% (data not shown). Ro20-1724 and papaverine have also been reported to inhibit the uptake of adenosine^{4,23}. Neither dipyrindamole nor a combination of 0.5

mM Ro20-1724, 0.5 mM papaverine and 0.1 mM dipyrindamole significantly reduced the effect of adenosine (data not shown). Accordingly, the effect cannot be due to an intracellular action of adenosine but must be mediated by extracellular receptors.

These receptors are different from the adrenergic α - and β -receptors present on the cells⁷ because: (1) methylxanthines antagonise the effect of adenosine (Fig. 1A, curve c, Fig. 1B, curve b) but not of α - or β -adrenergic agonists^{7,11}; (2) a 1,000-fold increase in the concentration of isoprenaline does not overcome the inhibition by adenosine¹¹; (3) the α -adrenergic antagonist phentolamine (100 μ M) does not alter the effect of adenosine¹¹. Thus, the receptors mediating the effect described here are probably specific for adenosine or a closely related compound.

The potencies of a number of derivatives of adenosine are summarised in Table 1. These results indicate that alterations in the ribose moiety of adenosine abolish the inhibitory potency of the nucleoside, whereas substituents in the purine moiety are comparatively well tolerated. This is consistent with the proposed¹ requirements of the R-site of adenylate cyclase. Adenosine not only inhibits but also stimulates the accumulation of cyclic AMP in the cultured brain cells^{11,24}. Also, this stimulatory effect is mediated by extracellular receptors which show the stereochemical requirements of an R-site. *N*⁶-(Phenylisopropyl)adenosine, however, a compound which is at least ten times more potent than adenosine in eliciting the inhibitory effect (Table 1), is only slightly less potent than adenosine in stimulating the accumulation of cyclic AMP in the cultures (our work, in preparation). Thus, the simplest explanation for these two effects of adenosine is the existence of two different types of receptor for the nucleoside. We propose the name A1-receptor for the one that mediates the inhibition and A2 for the one that mediates the stimulation of cyclic AMP accumulation.

The content of cyclic AMP in the incubation medium (about 20% of total) changed in proportion with the intracellular level of cyclic AMP. Thus, neither an increased release of cyclic AMP nor an increased degradation by PDE (Fig. 1A, curve b) can be responsible for the decrease in the intracellular accumulation of cyclic AMP. Thus the effect is likely to be due to inhibition of adenylate cyclase. This inhibition is probably a direct result of the binding of the A1-agonist to its receptor. Extracellular humoral factors distinct from adenosine are unlikely to mediate the effect. This was shown in the following way. After exposure to cells for 10 min, incubation medium containing 1 μ M isoprenaline and 1 μ M adenosine was treated with adenosine deaminase (Boehringer) to destroy adenosine. Cultures incubated with this medium revealed the same content of cyclic AMP as those exposed to control medium (only isoprenaline present) (data not shown).

Adenosine increases by way of extracellular receptors, the level of cyclic AMP in brain slices^{4,17} and in cultured cells of glial^{2,11} or neuronal^{19–21,23} origin. However, a receptor-mediated depression by adenosine of the level of cyclic AMP in brain slices has not been reported nor has it been described for any other system, with possibly one exception. In intact rat fat cells low concentrations of adenosine prevent the increase in the level of cyclic AMP that is evoked by noradrenaline^{25,26}. This effect is not blocked by dipyrindamole²⁶. Whether or not it is antagonised by methylxanthines remains to be clarified²⁷. Interestingly, *N*⁶-(phenylisopropyl)adenosine is also more potent in this system than adenosine²⁶. Glial precursor cells which constitute the majority of the cells in the cultures^{7,8} are probably responsible for the effects described. The presence of A1-receptors on cultured brain cells encourages investigations into the possibility that A1-receptors are also present on mature glial or neuronal cells in the brain.

The results presented here imply^{7,11} that A1-receptors and β -receptors are probably located on the same cells, most of which also display adrenergic α -receptors^{7,11}. Such a combination of receptors may be used as an identity label during isolation of the cells from heterogeneous populations.

Table 1 Potencies of derivatives of adenosine in inhibiting in cultured brain cells the increase in the level of cyclic AMP caused by isoprenaline

Compound	IC_{50} (μ M)
<i>N</i> ⁶ -(Phenylisopropyl)adenosine	0.005
2-Chloroadenosine	0.1
<i>N</i> ⁶ -Methyladenosine	0.5
Adenosine <i>N</i> ¹ -oxide	0.5
3'-Deoxyadenosine	5
2'-Deoxyadenosine	10
7-Deazaadenosine	10
9- β -D-Arabinofuranosyladenine	50
2', 3'-Isopropylideneadenosine	100

Experiments were performed with cells cultured for periods extending from 14 d to 23 d. The level of cyclic AMP measured after stimulation with isoprenaline (1 μ M) varied between 300 and 900 pmol per mg protein; 0.8–1.0 mg protein per plate. Other details as in Fig. 1.

This work was supported by the Sonderforschungsbereich 51 of the Deutsche Forschungsgemeinschaft. We thank Hoffmann-La Roche for Ro20-1724.

DIETRICH VAN CALKER
MARGARETE MÜLLER
BERND HAMPRECHT

Max-Planck Institut für Biochemie,
8033 Martinsried,
FRG

Received 18 July; accepted 24 October 1978.

1. Londos, C. & Wolff, J. *Proc. natn. Acad. Sci. U.S.A.* **74**, 5482-5486 (1977).
2. Clark, R. B., Gross, R., Su, Y.-F. & Perkins, J. P. *J. biol. Chem.* **249**, 5296-5303 (1974).
3. Haslam, R. J. & Rosson, G. M. *Molec. Pharmac.* **11**, 528-544 (1975).
4. Huang, M. & Daly, J. W. *Life Sci.* **14**, 489-503 (1974).
5. Clark, R. B. & Seney, M. N. *J. biol. Chem.* **251**, 4239-4246 (1976).
6. Sturgill, T. W., Schrier, B. K. & Gilman, A. G. *J. Cyclic Nucleotide Res.* **1**, 21-30 (1975).
7. Van Calker, D., Müller, M. & Hamprecht, B. *J. Neurochem.* **30**, 713-718 (1978).
8. Booher, J. & Sensenbrenner, M. *Neurobiology* **2**, 97-105 (1972).
9. Lim, R., Mitsunobu, K. & Li, W. K. P. *Expl. Cell Res.* **79**, 243-246 (1973).
10. Schousboe, A., Fosmark, H. & Svenneby, G. *Brain Res.* **116**, 158-164 (1976).
11. Van Calker, D. thesis, Univ. Munich (1977).
12. Laerum, O. D., Bigner, D. D. & Rajewski, M. F. (eds) *Biology of Brain Tumors Int. Un. Against Cancer Tech. Rep. Ser.* **30**, 143-157 (Geneva, 1978).
13. Bock, E., Möller, M., Nissen, C. & Sensenbrenner, M. *FEBS Lett.* **83**, 207-211 (1977).
14. Lim, R., Turriff, D. E., Troy, S. S., Moore, B. W. & Eng, L. F. *Science* **195**, 195-196 (1977).
15. Breen, G. A. M. & de Vellis, J. *Dev. Biol.* **41**, 255-266 (1974).
16. Brandt, M. *et al. Nature* **262**, 311-312 (1976).
17. Sattin, A. & Rall, T. W. *Molec. Pharmac.* **6**, 13-23 (1970).
18. Zenser, T. V. *Biochim. biophys. Acta.* **404**, 202-213 (1975).
19. Blume, A. J., Dalton, C. & Sheppard, H. *Proc. natn. Acad. Sci. U.S.A.* **70**, 3099-3102 (1973).
20. Penit, J., Huot, J. & Jard, S. *J. Neurochem.* **26**, 265-273 (1976).
21. Schultz, J. & Hamprecht, B. *Naunyn-Schmiedeberg's Arch. Pharmac.* **278**, 215-225 (1973).
22. Sheppard, H. & Wiggan, G. *Molec. Pharmac.* **7**, 111-115 (1970).
23. Green, R. D. & Stanberry, L. R. *Biochem. Pharmac.* **26**, 37-43 (1977).
24. Gilman, A. G. & Schrier, B. K. *Molec. Pharmac.* **8**, 410-416 (1972).
25. Fain, J. N., Pointer, R. H. & Ward, W. F. *J. biol. Chem.* **247**, 6866-6872 (1972).
26. Fain, J. N. *Molec. Pharmac.* **9**, 595-604 (1973).
27. Fain, J. N. & Wieser, P. B. *J. biol. Chem.* **250**, 1027-1034 (1975).
28. Lowry, O. H., Rosenbrough, N. J., Farr, A. L. & Randall, R. J. *J. biol. Chem.* **193**, 265-275 (1951).

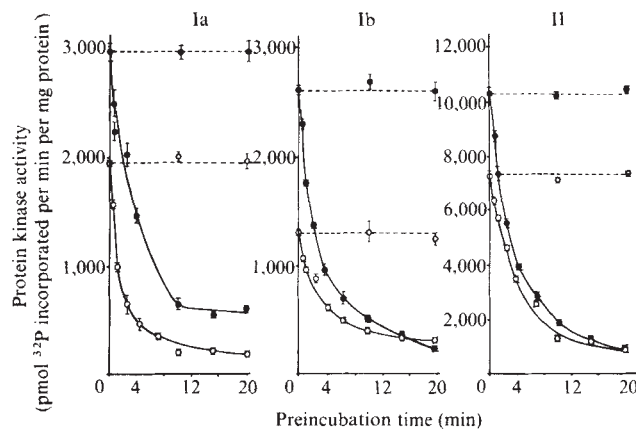


Fig. 1 Time-dependent blockade by indomethacin of types Ia, Ib and II cyclic AMP-dependent protein kinases. The standard assay mixture (final volume, 0.06 ml) for kinase activity contained: sodium α -glycerophosphate buffer (pH 7.5), 50 mM; dithiothreitol, 1 mM; protamine sulphate, 1.5 mg ml⁻¹; [γ -³²P]ATP, 6×10^{-4} M ($10 \mu\text{Ci} \mu\text{mol}^{-1}$); magnesium chloride, 10 mM; enzyme protein, 0.02–0.2 mg ml⁻¹; $\pm$ cyclic AMP, 2×10^{-5} M and indomethacin, 1×10^{-7} M (solubilised by adding a minimum quantity of 0.1 M NaOH to the aqueous suspension of the drug). Enzyme protein was added to assay mixture containing all components except Mg²⁺-³²P-ATP and preincubated for the indicated times at 30°C with (●—●) and without (○—○) cyclic AMP. The reaction was started by the addition of Mg²⁺-³²P-ATP. After incubation for 4 min at 30°C, the reaction was ended by the addition of 1 ml of cold 10% trichloroacetic acid (TCA) followed by bovine serum albumin, 100 μg , as a carrier. The protein precipitate was collected on Whatman GF/C class fibre filters, washed 5 times with 1-ml portions of cold 10% TCA, transferred to 10 ml of a Triton toluene-based scintillant, and counted in a liquid scintillation spectrometer. Protein concentration was determined by the method of Lowry *et al.*²¹ with bovine serum albumin as protein standard. Control values for protein kinase activity measured in the absence of indomethacin are depicted with (●—●) and without (○—○) cyclic AMP added. Each point represents the mean value $\pm$ s.e.m. obtained in 4 experiments carried out in a similar manner.

Indomethacin in submicromolar concentrations inhibits cyclic AMP-dependent protein kinase

INDOMETHACIN is a potent inhibitor of prostaglandin synthesis^{1,2}, and has been used to demonstrate indirectly the mediation of biological events by prostaglandins. The rationale for this use of indomethacin has been supported when analytical techniques for direct prostaglandin quantitation have been used^{3,4}. There are, however, conflicting data. A role for prostaglandins as the physiological messengers between hormones such as thyrotropin (TSH) and luteinising hormone (LH) and cyclic AMP has been proposed because prostaglandin antagonists inhibit the stimulatory effects of these hormones at their target tissues^{5,6}. A similar role has been proposed in the action of cholera toxin^{7,8}. In both cases, indomethacin failed to prevent the stimulatory effect of these agents on cyclic AMP production, giving support to the opposite view that prostaglandins were not essential intermediates⁹⁻¹¹. Paradoxically, indomethacin suppresses, *in vivo*, thyroid hormone secretion¹² and cholera toxin-induced intestinal fluid accumulation¹³⁻¹⁵. A post-cyclic AMP site of action for indomethacin may explain these discrepancies. To elucidate the role of cyclic AMP-

dependent protein kinase and endogenous protein phosphorylation in cholera toxin-induced intestinal secretion¹⁶, we have tested the influence of indomethacin on partially purified protein kinase from rabbit ileal mucosa. We report here that indomethacin in submicromolar concentrations inhibits cyclic AMP-dependent protein kinase and endogenous protein phosphorylation.

Homogenates of ileal mucosa were prepared by a modification of the method described previously¹⁷. The homogenisation was carried out in 10 mM α -glycerophosphate buffer, pH 7.5, containing 1 mM dithiothreitol. A 150,000g cytosol fraction was prepared by the method of Uno *et al.*¹⁸ and treated with solid ammonium sulphate followed by stirring for 30 min. Precipitation occurring between 30% and 55% saturation was removed by centrifugation at 20,000g for 20 min. The pellet was dissolved in and dialysed overnight against the α -glycerophosphate buffer. After dialysis the resulting solution was applied in a 2-ml volume to a column of DEAE-cellulose (1.5 $\times$ 60 cm) previously equilibrated with 10 mM α -glycerophosphate buffer (pH 7.5)/1 mM dithiothreitol. Protein kinases were eluted from

Table 1 Inhibition of rabbit ileal mucosa protein kinase activity and endogenous protein phosphorylation by indomethacin

Indomethacin effects	Protein kinase fractions					
	Ia		Ib		II	
Inhibition (ID ₅₀ , M)	+Cyclic AMP	-Cyclic AMP	+Cyclic AMP	-Cyclic AMP	+Cyclic AMP	-Cyclic AMP
Protein kinase activity	3.4×10^{-8}	1.7×10^{-8}	4.8×10^{-8}	1.7×10^{-8}	3.2×10^{-8}	1.8×10^{-8}
Endogenous protein phosphorylation	1.1×10^{-8}	0.60×10^{-8}	0.8×10^{-8}	0.62×10^{-8}	0.70×10^{-8}	0.58×10^{-8}
% Inhibition (1×10^{-7} M)	71 $\pm$ 2.1	84 $\pm$ 1.0	80 $\pm$ 1.2	88 $\pm$ 3.1	86 $\pm$ 1.0	87 $\pm$ 1.0

Values for % inhibition are mean $\pm$ s.e.m. of 4 determinations.

the column with a linear gradient (0–0.8 M) of NaCl. Three major cyclic AMP-dependent kinase activity peaks were resolved and arbitrarily designated types Ia, Ib and II. The peaks of protein kinase activity were eluted at the following approximate concentrations (in M) of NaCl: type Ia, 0.16; type Ib, 0.33; type II, 0.51. The active fractions under each peak were pooled and concentrated by ultrafiltration.

Our results demonstrate that indomethacin at submicromolar concentrations can inhibit the three active forms of cyclic AMP-dependent protein kinase resolved from rabbit ileal mucosa in a time-dependent and concentration-dependent manner. Indomethacin had no instantaneous effect on the rate of ^{32}P incorporation into protein substrate. After 4 min, the incubation time of the standard assay for protein kinase, the per cent inhibition ± 1 s.e.m. at 10^{-7} M indomethacin was (+cyclic AMP/–cyclic AMP) $17 \pm 3.7/20 \pm 4.3$ for peak Ia, $11 \pm 2.1/17 \pm 3.8$ for peak Ib and $13 \pm 2.9/16 \pm 6.2$ for peak II. Preincubation of the enzyme for different times before the addition of substrate led to progressive inhibition (Fig. 1). These results suggest that indomethacin takes part in a time-dependent reaction before it can produce its maximal inhibitory effect. The inhibition of cyclic AMP-dependent protein kinase activity by indomethacin was concentration-dependent (Fig. 2), and could not be reversed by dilution of the inhibitor, suggesting it was either irreversible or a result of tight binding by indomethacin (data not shown). Putative phosphorylation of endogenous protein substrate was measured by the incorporation of ^{32}P from ^{32}P -ATP to trichloroacetic acid-insoluble material (without exogenous protamine as substrate) of each activity peak. The inhibition response was qualitatively similar to the results obtained when exogenous substrate is added (Table 1).

The ID_{50} (concentration of inhibitor required to inhibit 50% of enzyme activity)¹⁹ and per cent inhibition are shown in Table 1. The ID_{50} for indomethacin, when computed for protein kinase activity, ranged from 1.7×10^{-8} M (–cyclic AMP) to 4.8×10^{-8} M (+cyclic AMP). The ID_{50} for endogenous protein phosphorylation ranged from 0.58×10^{-8} M (–cyclic AMP) to 1.1×10^{-8} M (+cyclic AMP). The per cent inhibition for protein kinase activity at 1×10^{-7} M indomethacin varied from 71% to 88%.

In addition to inhibiting prostaglandin synthetase, indomethacin exerts effects on a variety of enzyme systems at micromolar concentrations or greater³, properties which should be kept in mind when using this drug to evaluate the role of prostaglandins in biological events. On a molar basis indomethacin was between 79 and 224 times more potent as an inhibitor of cyclic AMP-dependent protein kinase activity than

of prostaglandin synthetase activity (rabbit kidney microsomes³). Furthermore, at normal therapeutic doses, the free plasma concentration of indomethacin (5.6×10^{-7} M)⁴ is 12–33 times greater than the ID_{50} for protein kinase.

This study demonstrates that indomethacin is a potent inhibitor of cyclic AMP-dependent protein kinases and endogenous phosphorylation. It has been suggested that activation of protein kinase represents a major mechanism by which cyclic AMP carries out its function as an intracellular messenger in mammalian systems²⁰. These data suggest that at therapeutic doses, indomethacin may function as an antagonist to cyclic AMP. Further studies are necessary to identify the mechanism of inhibition and to establish whether this new action of indomethacin occurs more generally in different tissues.

This work was supported by the Medical Research Service of the Veterans Administration.

HARVEY S. KANTOR
MICHAEL HAMPTON

Division of Infectious Diseases,
Department of Medicine,
University of Health Sciences,
The Chicago Medical School,
North Chicago, Illinois 60064
and
V.A. Hospital, North Chicago, Illinois

Received 13 June; accepted 6 November 1978.

1. Vane, J. R. *Nature new Biol.* **231**, 232–235 (1971).
2. Flower, R. J. & Vane, J. R. *Biochem. Pharmacol.* **23**, 1439–1449 (1974).
3. Flowers, R. J. *Pharmac. Rev.* **26**, 33–67 (1974).
4. Ferreira, S. H. & Vane, J. R. *A. Rev. Pharmacol.* **14**, 57–73 (1974).
5. Yu, S.-C., Chang, L. & Burke, G. J. *clin. Invest.* **51**, 1038–1042 (1972).
6. Kuehl, F. A., Humes, J. L., Tarnoff, J., Cirillo, V. J. & Ham, E. A. *Science* **169**, 883–886 (1970).
7. Bennett, A. *Nature* **231**, 536 (1971).
8. Vaisrub, S. J. *Am. med. Ass.* **219**, 213 (1972).
9. Spaulding, S. W. & Burrow, G. N. *Endocrinology* **96**, 1018–1021 (1975).
10. Zor, U. *et al. Prostaglandins* **4**, 499–507 (1973).
11. Kimberg, D. V., Field, M., Gershon, E. & Henderson, A. J. *clin. Invest.* **53**, 941–949 (1974).
12. Thompson, M. E., Orczyk, G. P. & Hedge, G. A. *Endocrinology* **100**, 1060–1067 (1977).
13. Jacoby, H. I. & Marshall, C. H. *Nature* **235**, 163–165 (1972).
14. Gots, R. E., Forman, S. B. & Gianella, R. A. *J. infect. Dis.* **130**, 280–284 (1974).
15. Wald, A., Gotterer, G. S., Rajendra, G. R., Turjman, N. A. & Hendrix, T. R. *Gastroenterology* **72**, 106–110 (1977).
16. Kantor, H. S. & Hampton, M. *Clin. Res.* **25**, 587A (1977).
17. Kantor, H. S., Tao, P. & Gorbach, S. L. *J. infect. Dis.* **129**, 1–9 (1974).
18. Uno, I., Ueda, T. & Greengard, P. *J. biol. Chem.* **251**, 2192–2195 (1976).
19. Job, D., Coheti, C., Dhien, A. & Chambaz, E. M. *Analyt. Biochem.* **84**, 68–77 (1978).
20. Kuo, J. F. & Greengard, P. *Proc. natn. Acad. Sci. U.S.A.* **64**, 1349–1355 (1969).
21. Lowry, O. H., Rosebrough, N. J., Farr, A. L. & Randall, R. J. *J. biol. Chem.* **193**, 265–275 (1951).

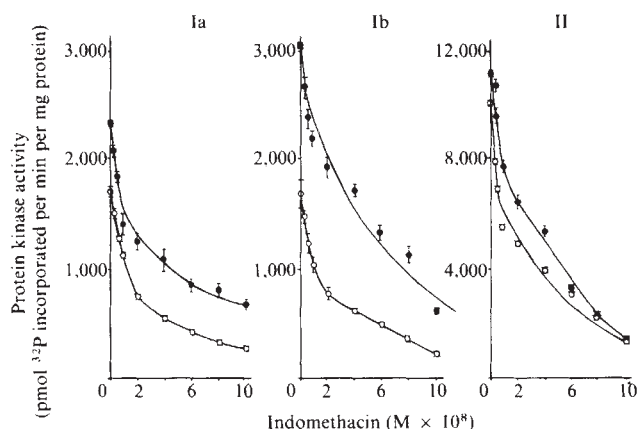


Fig. 2 Concentration-dependent inhibition by indomethacin of types Ia, Ib and II cyclic AMP-dependent protein kinase activities. Enzyme protein was added to the assay mixture containing the indicated concentration of indomethacin and preincubated for 20 min before starting the reaction by the addition of the substrate, Mg^{2+} - ^{32}P -ATP. Each point, with (●—●) and without (○—○) cyclic AMP, represents the mean value $\pm$ s.e.m. obtained in 4 experiments carried out in a similar manner.

Agropine—a major new plasmid-determined metabolite in crown gall tumours

THE crown gall tumour system may represent an elegant form of biochemical parasitism, in which the tumour-inducing bacterium *Agrobacterium tumefaciens* can, by insertion¹ and expression² of genetic sequences carried on the tumour-inducing (Ti) plasmid³, determine the synthesis of unusual amino acids in transformed plant cells. These abnormal metabolites can be specifically utilised as nitrogen sources by the particular strain of inducing bacterium⁴. So far, two classes of plasmid-determined product have been recognised: the octopine family—octopine⁵, octopinic acid⁶, lysopine⁷ and histopine⁸—and the nopaline family—nopaline⁹ and nopalinic acid¹⁰. Whether the octopine or nopaline family of amino acids is synthesised by a particular tumour depends entirely on the type of Ti plasmid present in the inducing bacterium¹¹. We report

Table 1 Octopine, nopaline and agropine degradation by *A. tumefaciens* strains and synthesis by crown gall tumours induced by those strains

Bacterial strain	Bacterial degradation*			Synthesis in tumour†		
	Octopine	Nopaline	Agropine	Octopine	Nopaline	Agropine
A6	+	-	+	+	-	+
A66	+	-	+	+	-	+
B6	+	-	+	+	-	+
181	-	+	-	-	-	-
38-9	-	+	-	-	+	-
T-37	-	+	-	-	+	-
C-58	-	+	-	-	+	-

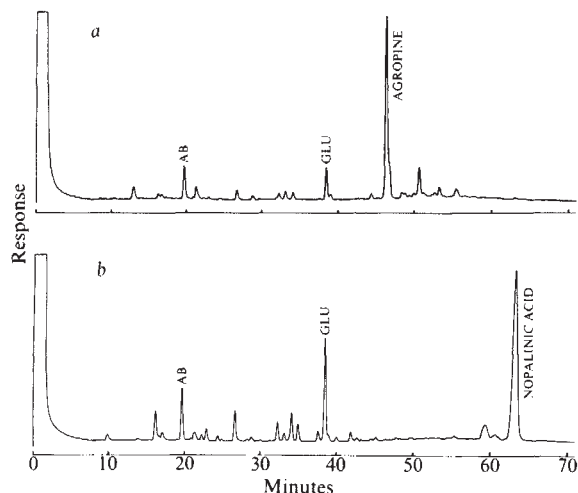
* Bacteria were inoculated into a liquid medium¹⁶ containing either glucose (0.4% w/v) as C and energy source, and octopine (0.1% w/v) or nopaline (0.1% w/v) as sole N sources or agropine (0.1% w/v) as sole carbon and energy source, with ammonium sulphate (0.2% w/v) as N source. Agropine was isolated from an aqueous extract of A66-induced tumour tissue by adsorption on to Dowex 50 × 8, 20-50 mesh, H⁺ form. The resin was washed with 10 bed volumes of water and eluted with 0.1 M ammonium hydroxide. Fractions equal to the column bed volume were collected, and those showing strong reaction with the alkaline silver nitrate reagent¹³ after paper chromatography in propan-1-ol-ammonia (3:2 by volume) were combined and evaporated at 50 °C. When required, further purification was obtained by preparative paper chromatography. Cultures were shaken at 28 °C, and were examined for growth visually after 24 h. Disappearance of octopine or nopaline (measured colorimetrically using the α -naphthol-diacyl reagent¹⁷) or disappearance of agropine (measured by gas chromatography¹²) correlated with visible bacterial growth.

† Synthesis of octopine, nopaline or agropine by bacteria-free tumour tissue was demonstrated by incubating the tumour tissue (~50 mg fresh weight) with either [U-¹⁴C]arginine or [U-¹⁴C]D-glucose (1 μ Ci, 304 μ Ci μ mol⁻¹) for 24 h at 20 °C. The tissue was freeze dried and after grinding with sand, extracted with 2 × 5 ml water. After centrifugation the supernatant was passed through a column of Zerotit 225, 52-100 mesh, H⁺ form, 1 ml bed volume. The resin was washed with water and cationic compounds (including octopine, nopaline and agropine) were eluted with 2M ammonium hydroxide. This eluate was concentrated by evaporation, and the individual compounds present were separated by two dimensional paper (Whatman 3 MM) chromatography (solvent I, 100 g phenol + 39 ml water; solvent II, butan-1-ol-propionic acid water 92:47:61), with added unlabelled octopine, nopaline and agropine. These compounds were located using either the Sakaguchi reaction¹⁸ (octopine and nopaline) or the alkaline silver nitrate reagent¹³, and the ¹⁴C present in each compound was assayed by liquid scintillation. Radioactivity present in the agropine spot in '+' samples ranged from 1.7 × 10⁴ c.p.m. (A66) to 3.4 × 10⁴ c.p.m. (B6); '-' indicates c.p.m. < 10². Lower limit of detectability was 10² c.p.m. No label was detected in agropine (added as a 'cold' standard before extraction) when healthy *N. tabacum* was incubated with ¹⁴C-glucose as described above. R_f values for agropine were: solvent I 0.78; solvent II 0.25.

here the discovery of a completely new type of plasmid-determined compound which is present in prodigious amounts (up to 7% of the tumour dry weight) in tumours induced by octopine strains of *A. tumefaciens*.

Gas chromatograms of the heptafluorobutyric (HFB) *n*-propyl derivatives¹² of the free amino acids isolated from bacteria-free crown gall tumours initiated on 'White Burley' tobacco

Fig. 1 Gas chromatograms of the HFB *n*-propyl derivatives¹² of the free amino acids extracted (see Table 1) from crown gall tumour tissue induced on *Nicotiana tabacum* by *A. tumefaciens* strain A66(a) and strain T37(b). AB, α -aminobutyric acid (added as internal standard); GLU, glutamic acid. Nopaline, although present in the T37 tumour at approximately the same level as nopalinic acid, does not chromatograph in these conditions.



plants¹⁰ by inoculation with any one of a series of octopine-metabolising strains of *A. tumefaciens* (A6, B6, A66) showed the presence of a major unidentified peak (Fig. 1a). This peak, which we have named 'agropine', was not detected either in tumours induced by nopaline-metabolising strains (Fig. 1b) of *A. tumefaciens* (C58, T-37, 38-9, 181) or in healthy tobacco tissue.

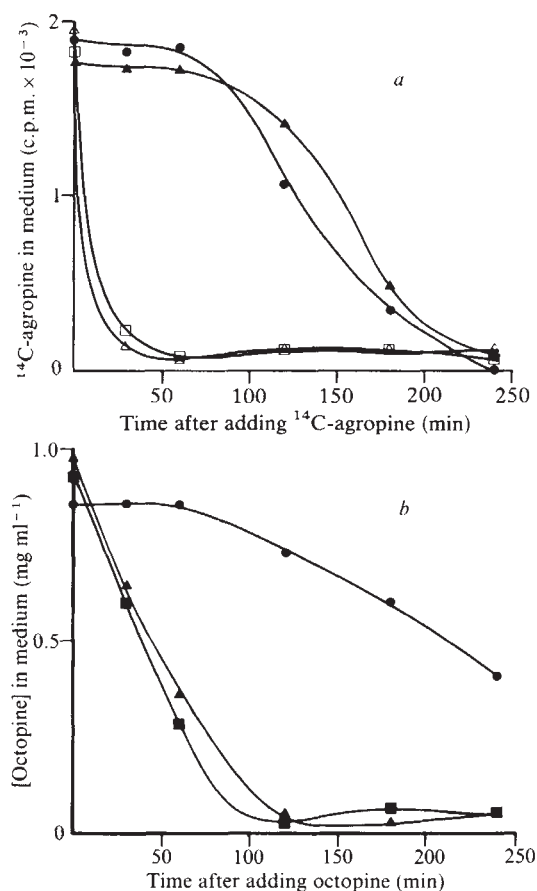


Fig. 2 Uptake of ¹⁴C-agropine and octopine by *A. tumefaciens* strains. Bacteria were cultured at 28 °C in a liquid minimal medium¹⁶ containing either glucose (0.4% w/v) + ammonium sulphate (0.2%), glucose (0.4%) + octopine (0.1%) or agropine (0.2%) + ammonium sulphate (0.2%). The bacteria were collected by centrifugation after 24 h, washed in minimal medium with no added carbon or nitrogen sources, and then resuspended in sufficient fresh minimal medium to give a final A₆₃₀ value of 1.5. After addition of ¹⁴C-agropine (~2 × 10⁵ c.p.m.), isolated from A66-induced tumour tissue after incubation with [U-¹⁴C]glucose (see Table 1) or octopine (1 mg) the culture was shaken at 28 °C, and samples (10%) of the medium assayed, after centrifugation, by liquid scintillation (¹⁴C-agropine uptake) or by the α -naphthol-diacyl reagent¹⁷ (octopine uptake). ●, Strain B6, not induced; ■, B6, induced with octopine; □, B6, induced with agropine; ▲, strain 1-1, not induced; △, 1-1, induced with agropine.

Transfer of the Ti plasmid from an octopine strain (B6S3) by conjugation in the presence of octopine (2 mg ml⁻¹), to a plasmid-cured rifampicin-resistant mutant of the nopaline strain C58, resulted in transconjugants (selected by plating on a minimal medium containing octopine (2 mg ml⁻¹) and rifampicin 25 μ g ml⁻¹) which induced tumours containing agropine and octopine. These data suggest that agropine, like members of the octopine and nopaline families, is a strain-specific, plasmid-determined product.

The new compound was isolated (Table 1) from bacteria-free tumour tissue as colourless microcrystals from acetone-water, melting point 129 °C, [α]_D²⁵ +19.5° (*c* = 1 in water). High resolution mass spectrometry indicated a molecular weight of 275.10254, corresponding to a molecular formula C₁₁H₁₇NO₇. Spectroscopic studies (IR, NMR) suggested a polyhydroxylic carboxylic acid. Gas chromatography of the HFB *n*-propyl

derivative of purified agropine gave a single peak which co-chromatographed with the large peak found in the derivatised amino acid fraction from the tumour tissue. The molecule does not react with ninhydrin, but can be detected on paper using the alkaline silver nitrate reagent¹³.

All octopine strains of *A. tumefaciens* (A6, B6, A66) examined so far utilise agropine as a sole carbon and energy source (Table 1). Uptake of agropine seems to be mediated by an inducible permease (Fig. 2a), and results obtained with the mutant octopine strain 1-1, which is constitutive for octopine uptake (Fig. 2b) and degradation, but not for agropine uptake (Fig. 2a) suggest that the uptake of octopine and agropine is controlled by different genes. Nopaline strains do not take up agropine after pregrowth in a minimal medium either with or without added octopine, nopaline or agropine, although these strains do transport and degrade octopine after pregrowth with nopaline.

The evidence suggests that agropine may represent a condensation product between a sugar or sugar derivative and an amino acid, and is therefore probably not synthesised by the enzyme responsible for synthesis of the octopine family of amino acids. Estimates of the amount of plasmid-derived genetic material in the tumour genome suggests that information sufficient to code for only four proteins of molecular weight ~50,000 is present¹⁴. Thus it is unlikely that agropine synthesis in the tumour could involve many intermediates not normally present in the plant tissue before transformation.

There is little doubt that agropine is not a normal plant product, although it is still conceivable that the capacity to synthesise agropine resides in the plant genome, and is only expressed in response to infection by a particular strain of infecting bacterium. However, this hypothesis is extremely unlikely since agropine is present in tumours induced on several unrelated genera of host plant¹⁵.

In any event, the remarkably high levels of agropine present in tumours represent a major diversion of the plants' nutritional resources into a form which is specifically available as a carbon and energy source to the tumour-inducing pathogen. Agropine, together with the specific nitrogen sources octopine, octopinic acid, lysopine and histopine, could therefore provide the basic nutritional requirements of the pathogen in a form available only to that strain.

We thank Dr A. G. Hepburn, University of Bristol, for carrying out the bacterial conjugation.

J. L. FIRMIN

John Innes Institute,
Colney Lane, Norwich, UK

G. R. FENWICK

Food Research Institute,
Colney Lane, Norwich, UK

Received 3 July; accepted 3 November 1978.

- Chilton, M.-D. *et al.* *Cell* **11**, 263-271 (1977).
- Drummond, M. H., Gordon, M. P., Nester, E. W. & Chilton, M.-D. *Nature* **269**, 535-536 (1977).
- Watson, B., Currier, T. C., Gordon, M. P., Chilton, M.-D. & Nester, E. W. *J. Bact.* **123**, 255-264 (1975).
- Bomhoff, G. thesis, Univ. Leiden (1974).
- Ménagé, A. & Morel, G. *C. r. heb. Séanc. Acad. Sci. Paris* **259**, 4759-4796 (1964).
- Ménagé, A. & Morel, G. *C. r. heb. Séanc. Acad. Sci. Paris* **261**, 2001-2002 (1965).
- Biemann, K., Lioret, C., Asselineau, J., Lederer, E. & Polonsky, J. J. *Bull. Soc. Chim. biol.* **42**, 979-991 (1960).
- Kemp, J. D. *Biochem. biophys. Res. Commun.* **74**, 862-868 (1977).
- Goldmann, A., Thomas, D. W. & Morel, G. *C. r. heb. Séanc. Acad. Sci. Paris* **268**, 852-854 (1969).
- Firmin, J. L. & Fenwick, R. G. *Phytochemistry* **16**, 761-762 (1977).
- Bomhoff, G., *et al.* *Molec. gen. Genet.* **145**, 177-181 (1976).
- March, J. F. *Analyt. Biochem.* **69**, 420-442 (1975).
- Trevelyan, W. E., Procter, D. P. & Harrison, J. S. *Nature* **166**, 444-445 (1950).
- Hack, E. & Kemp, J. D. *Biochem. biophys. Res. Commun.* **78**, 785-791 (1977).
- Scott, I. M. *et al.* (in preparation).
- Zevenhuizer, L. P. T. M. *J. gen. Microbiol.* **68**, 239-243 (1971).
- Johnson, R. *et al.* *Proc. natn. Acad. Sci. U.S.A.* **71**, 536-539 (1974).
- Jepson, J. B. & Smith, I. *Nature* **172**, 1100-1101 (1953).

A physically altered leucyl-tRNA synthetase complex in a CHO cell mutant

CONDITIONALLY lethal mutants in animal cells have been of great value in advancing our knowledge of genetics, biochemistry and molecular biology. Mammalian cells with temperature sensitive (*ts*) mutant phenotypes have been isolated by several laboratories, however no thorough characterisations of the altered gene products are available¹. Selection procedures have been devised by Thompson *et al.* to isolate *ts* aminoacyl-tRNA synthetase variants in Chinese hamster ovary (CHO) cells which behave as missense mutants^{2,3}. The CHO temperature-sensitive mutant *tsH1* has very low *in vivo* levels of leucyl-tRNA, compared to wild-type levels at non-permissive temperatures; levels of the other 19 aminoacyl-tRNAs are unchanged⁴. This mutant also has an altered leucyl-tRNA synthetase which is thermosensitive *in vitro*⁴, enzymatically different from wild-type leucyl-tRNA synthetase⁵, and has an altered *K_m* for leucine⁵. We have now shown the leucyl-tRNA synthetase in *tsH1* also has altered physical characteristics. The mutant enzyme exists primarily in a low molecular weight form whereas wild-type leucyl-tRNA synthetase is primarily associated with higher molecular weight particulate forms.

The particulate nature of aminoacyl-tRNA synthetases is a common feature of eukaryotic cells. The nature of these particulate enzyme forms reflects *in vivo* molecular organisation but its significance is not clear⁶. The size and form distributions of the 13 post-ribosomal particulate synthetases examined were indistinguishable in wild-type and mutant cells, except for leucyl-tRNA synthetase. The dramatic physical alteration of leucyl-tRNA synthetase is shown in Fig. 1. The distributions of

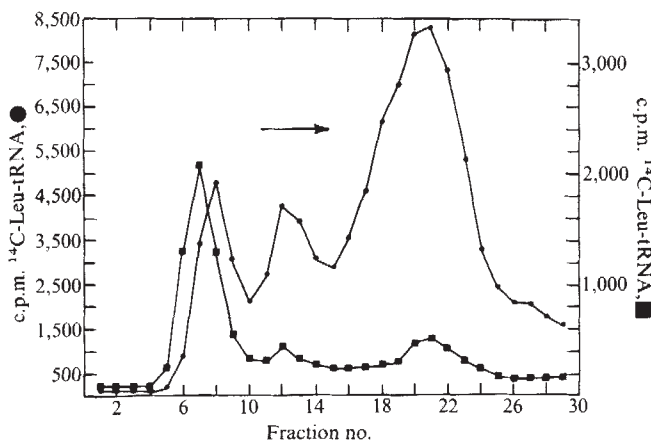


Fig. 1 Size distribution of leucyl-tRNA synthetase on sucrose gradients. Enzyme from wild type (●) and mutant *tsH1* (■) Chinese hamster ovary cells. Cells were grown in suspension culture at 34.5 °C in α -minimal essential medium¹¹ (Flow) containing 15% calf serum with penicillin (100 IU ml⁻¹) and streptomycin (100 μ g ml⁻¹). Cell doubling time was 16 h for wild-type cells and 21 h for *tsH1*. Cells were collected during exponential growth, washed with 0.25 M sucrose, suspended in buffer A (100 mM KCl, 10 mM Tris-HCl (pH 7.5 at 25 °C), 1.5 mM MgCl₂, and 0.1 mM dithiothreitol) (18 ml buffer per 10⁹ cells) and stored at -90 °C. Cells (10⁹) were broken in buffer A (20 ml) containing 1% Nonidet P-40 (Particle Data Labs) and supernatant was centrifuged at high speed to remove ribosomes. This postribosomal supernatant was centrifuged further to give a 6-40S subcellular pellet which was resuspended in buffer B (10 mM KCl, 10 mM Tris-HCl (pH 7.5 at 25 °C), 1.0 mM MgCl₂, and 0.1 mM dithiothreitol), layered on 10-30% w/v logarithmic sucrose gradients in buffer B, and centrifuged for 22 h at 25,000 r.p.m. in a Spinco SW-25.2 rotor at 4 °C (refs 12, 13). Sucrose gradients were assayed for the amino acids Arg, Asp, Cys, Gln, Glu, His, Ile, Leu, Lys, Met, Pro, Thr, and Val by esterifying ¹⁴C-amino acids to cognate rat liver tRNAs and determining products formed (¹⁴C-aminoacyl-tRNAs)^{13,14}.

the 12 indistinguishable activities are not shown here, as similar wild-type profiles have been previously published⁶.

The wild-type leucyl-tRNA synthetase shows three particulate forms: an 8S form, a 20S form and a 30S form. Most of the wild-type enzyme activity is reproducibly found in the large 30S particle, and the mutant leucyl-tRNA synthetase is predominantly found in the small 8S form. The mutant leucyl-tRNA synthetase maintains negligible amounts of the higher molecular weight forms. These differences in synthetase forms provide additional evidence that the cause of the *tsH1* phenotype is a directly altered leucyl-tRNA synthetase. These differences alone do not exclude the possibility that the shifts in leucyl-tRNA synthetase are due to a change in the molecules with which the enzyme is complexed, rather than a change in the leucyl-tRNA synthetase itself. However, previous studies on the mutant cells show this to be unlikely^{4,5}.

The physical evidence for a structurally altered leucyl-tRNA synthetase further supports the original evidence of Thompson *et al.*⁴ that *tsH1* is a true genotypic variant with an altered gene product and further counters epigenetic theories that somatic cell variants have stable phenotypes but do not contain changes in DNA structure. The electrophoretically altered hypoxanthine-guanine phosphoribosyltransferase from mutant HeLa cells⁷⁻⁹ and cyclic AMP-dependent protein kinase from mutant S49 lymphoma cells¹⁰ represent strong physical evidence for the existence of directly altered gene products and, consequently, stable genetic variants. The fact that the *tsH1* gene product, leucyl-tRNA synthetase, is temperature-sensitive *in vitro*, and is apparently physically altered as well, suggests that it is also structurally altered due to a missense mutation in the structural gene.

The difference in synthetase forms between this mutant and wild-type cells is also very good evidence that the particulate nature of aminoacyl-tRNA synthetases is not an artefact but indeed a real intramolecular assemblage. The fact that large particulate forms of leucyl-tRNA synthetase are largely absent in the mutant although the mutant still survives suggests the complex is not completely essential for cellular function in the conditions described in this report. This mutant should prove very useful in determining the cellular role of the particulate forms of this enzyme.

We thank Bruce Ruefer, Helen Barrington and John Hanners for technical assistance. This work was supported by NIH grant GM 19506 and RCDA to A.H. and USERDA and EPA agreement EPA-IAG-D5-E681 to M.D.E. and P.R.

ARNOLD E. HAMPEL

Northern Illinois University,
Department of Biological Sciences,
DeKalb, Illinois 60115

PRESTON O. RITTER

Eastern Washington University,
Department of Chemistry,
Cheney, Washington 99004

M. DUANE ENGER

Los Alamos Scientific Laboratory,
University of California,
Los Alamos, New Mexico 87545

Received 14 August; accepted 3 November 1978.

1. Basilico, C. *Cancer Res.* **24**, 222-266 (1977).
2. Thompson, L. H., Stanners, C. P. & Siminovich, L. *Somatic Cell Genet.* **1**, 189-208 (1975).
3. Thompson, L. H., Lofgren, D. J. & Adair, G. M. *Cell* **11**, 157-168 (1977).
4. Thompson, L. H., Harkins, J. L. & Stanners, C. P. *Proc. natn. Acad. Sci. U.S.A.* **70**, 3094-3098 (1973).
5. Haars, L., Hampel, A. & Thompson, L. H. *Biochim. biophys. Acta* **454**, 493-503 (1976).
6. Ritter, P., Enger, M. D. & Hampel, A. *Onco-Developmental Gene Expression* (eds Fishman, W. H. & Sell, S.) 47-56 (Academic, New York, 1976).
7. Milman, G., Lee, E., Ghangas, G. S., McLaughlin, J. R., & George, M., Jr *Proc. natn. Acad. Sci. U.S.A.* **73**, 4589-4593 (1976).
8. Milman, G., Krauss, S. & Olsen, A. *Proc. natn. Acad. Sci. U.S.A.* **74**, 926-930 (1977).
9. Fenwick, R. G., Jr, Sawyer, T. H., Kruh, G. D., Astrin, K. H. & Caskey, C. T. *Cell* **12**, 383-391 (1977).
10. Steinberg, R., O'Farrell, P. H., Friedrich, V. & Coffino, P. *Cell* **10**, 381-391 (1977).
11. Stanners, C. P., Elicieri, G. L. & Green, H. *Nature new Biol.* **230**, 52-54 (1971).
12. Enger, M. D., Walters, R. A., Hampel, A. & Campbell, E. W. *Eur. J. Biochem.* **43**, 17-28 (1974).
13. Hampel, A. E., Enger, M. D. & Ritter, P. *Meth. Enzym.* **59**, 229-234 (1978).
14. Hampel, A. & Enger, M. *J. molec. Biol.* **79**, 285-293 (1973).

Nucleotide sequence at the insertion sites of a kanamycin transposon

THE TRANSPOSON Tn903 carrying a gene for kanamycin resistance, is 3,100 base pairs in size and contains an inverted repeat sequence of 1,050 base pairs at both ends¹⁻⁴. We report here the generation of a 9-base pair repeated sequence at the insertion site of Tn903. Tn903 can be transposed to at least nine different sites on the coliphage fd DNA^{1,2}. We have also transposed Tn903 to three sites on small colicin E1 plasmid (ColE1) derivatives of about 1,600 base pairs. The presence of these many insertion sites on small recipient DNA molecules implies that no specific sequence is involved in the target sites on the recipient. To investigate the mechanism of transposition of Tn903 at the molecular level, we have now analysed three independent insertions of Tn903 on the small ColE1 DNA molecules. The nucleotide sequences of the three target sites and of the corresponding junctions between Tn903 and the small ColE1 have been determined.

The small ColE1 derivatives used were pAO2 (ref. 5) and pAO3. pAO2 is constructed from the region between 0.75 and 1.00 map units of the ColE1 DNA molecule. (A map unit is the fractional length from the 5' end of the H-strand⁶ generated from ColE1 DNA by cleavage with the restriction endonuclease *EcoRI*.) pAO3 is similar to pAO2 but carries an additional region of 70 base pairs (Fig. 1). pAO43 (pAO2::Tn903), pAO48 (pAO2::Tn903) and pAO65 (pAO3::Tn903) were chimaeric plasmids carrying Tn903, constructed by transposing Tn903 from R6-5 to either pAO2 or pAO3. These chimaeras contained a complete set of Tn903, demonstrated by their ability to donate the kanamycin marker to phage fd and λ b5. Restriction endonucleases used were *HaeIII*, *HapII*, *HgaI*, *HhaI*, *BpaI*, *BpaII*, *HindIII* and *EcoRI*. Methods for preparation of enzymes, plasmid DNA and restriction fragments have been described earlier^{5,7}. DNA sequencing was carried out as reported by Maxam and Gilbert⁸.

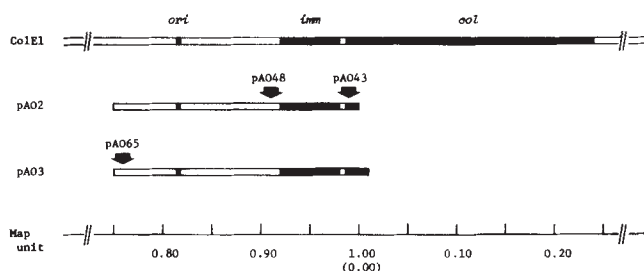


Fig. 1 Insertion sites of Tn903 on the small ColE1 DNA molecules. The Tn903 insertion sites in pAO43, pAO48 and pAO65 are shown by an arrow on the genetic map of pAO2 and pAO3. At the top is part of the genetic map of ColE1 as reference. DNA replication initiates at *ori* and proceeds towards the left. *col* is the structural gene of colicin E1 protein and *imm* is a gene conferring immunity to colicin E1 upon the host cell. The map position is indicated by the map unit of ColE1.

The Tn903 insertion sites in pAO43, pAO48 and pAO65 were determined by comparison of restriction fragments produced from these plasmids with those from pAO2 or pAO3 by digestion with various restriction endonucleases (data not shown), and the sites were assigned to be 0.99, 0.91 and 0.76 map units on the small ColE1 DNA molecules, respectively (Fig. 1). The cleavage maps⁵ around the three insertion sites and restriction fragments used for sequence analysis are shown in Fig. 2. The sequences within and surrounding the three insertion sites deduced are presented in Fig. 3a-c, in comparison with those in the corresponding region of pAO2 or pAO3. There are two particularly noteworthy features of the sequence. First, the two ends of Tn903 are identical for at least 250 base pairs from the insertion site (Fig. 3d), in agreement with previous

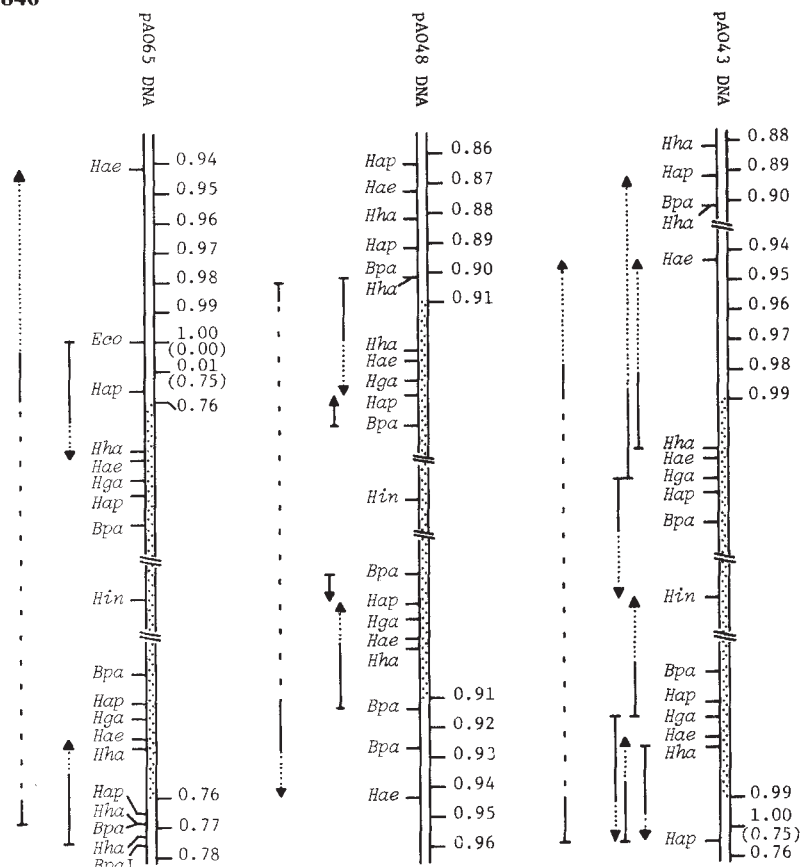


Fig. 2 Cleavage maps around the three insertion sites of Tn903 and labelled strands used for sequence analysis. The map position is indicated by the map unit of ColE1 on each plasmid DNA. Cleavage sites by restriction endonucleases are shown under plasmid DNA: *Hae*, *Hap*, *Hha*, *Hga*, *Bpa*, *Hin* and *Eco* represent *Hae*III, *Hap*II, *Hha*I, *Bpa*II, *Hin*III and *Eco*RI, respectively. Arrows indicate [5'-³²P]labelled strands aligned in the 5' to 3' direction. The bottom arrow in each plasmid is a strand prepared from pA02 or pA03 DNA. All the strands were obtained by secondary cleavage of [5'-³²P]labelled double-stranded DNA. Solid lines in the arrows indicate the sequences that were resolved, and the dotted portions of the arrows indicate the sequences that were not as well resolved. The shading represents the Tn903 DNA.

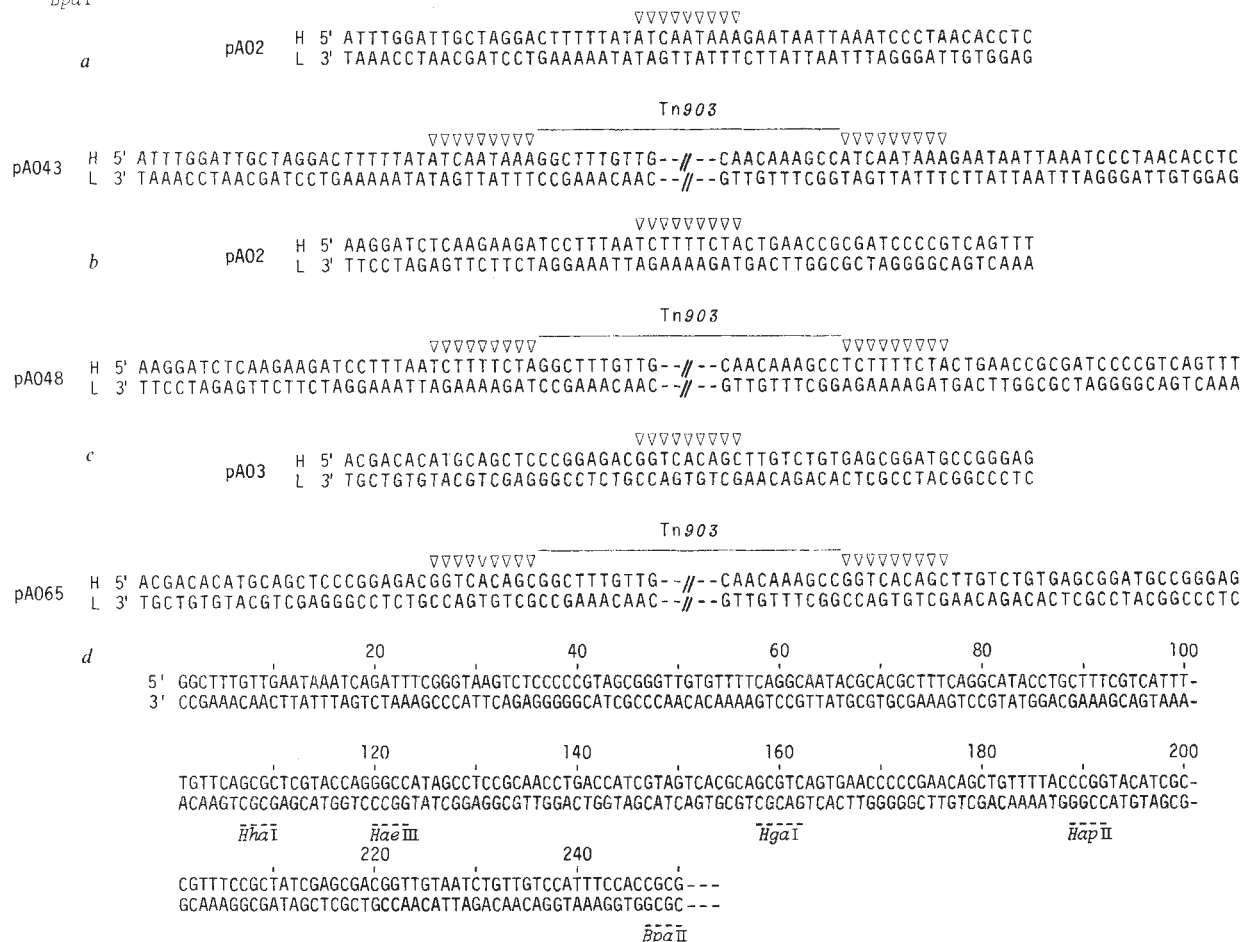


Fig. 3 The nucleotide sequence surrounding the Tn903 insertion sites. The upper part in *a*, *b* and *c* represents the nucleotide sequence of the pA02 or pA03 DNA molecule, surrounding the Tn903 insertion sites. The lower part in *a*, *b* and *c* shows the nucleotide sequence covering the junction between Tn903 and either pA02 or pA03 in pA043, pA048 and pA065, respectively. ▽ shows the duplication of 9 base pairs on Tn903 insertion. *d*, The nucleotide sequence of the inverted repeat carried by Tn903. This sequence represents approximately 25% of the inverted repeat sequence. Nucleotide position is numbered from the integration site.

heteroduplex analyses^{1,2,4}. This sequence has no obvious similarities to those of IS1 and IS2 (refs 9–12). Second, a 9-base pair segment, present at the site of insertion on the recipient DNA molecule, is repeated in the same orientation at both ends of Tn903. It is unlikely that the 9-base pair segment was brought from R6–5 as the donor of Tn903, because the sequence of 9-base pair repeat is different for each insertion analysed. Similar repetition has also been observed with insertion of IS1 (refs 9, 10), Tn3 (E. Ohtsubo, personal communication) and Tn5 (H. Schaller, personal communication). Therefore, it seems that this duplication is a general feature of the Tn and IS translocation process, suggesting all Tn and IS use a similar mechanism of transposition.

The creation of a 9-base pair can be explained as follows. Two single-strand cuts separated by 9-base pairs are introduced on opposite strands of the recipient DNA molecule. The two new 5' ends formed at the insertion site are then joined to the 3' termini of Tn903 so as to conserve the nine bases of the recipient DNA at both sides of the Tn sequence. The unpaired gaps at the complementary strands are then filled by repair synthesis. A similar explanation has been discussed in detail by Caros *et al.*⁹ and Grindley¹⁰.

Comparison of the three target sites indicated that neither the duplicated regions nor their vicinity contained an unique sequence common to three insertions. This suggests that no recognition of a specific sequence (recognition sequence) is involved for Tn903 translocation, and that Tn903 can be transposed to any site on DNA. However, we cannot rule out the possibility that the recognition sequence occurs very far from the target site for insertion.

We thank A. Maxam and W. Gilbert for providing the details of the procedure for determining DNA sequences before publication, S. Cohen, P. Kretschmer and H. Yamagishi for bacterial and phage strains, and E. Ohtsubo and H. Schaller for sending us their results before publication. This work was supported by Grant-in-Aid for Scientific Research from the Ministry of Education, Science and Culture, Japan.

ATSUHIRO OKA
NOBUO NOMURA
KAZUNORI SUGIMOTO
HIROYUKI SUGISAKI
MITURU TAKANAMI

*Institute for Chemical Research,
Kyoto University, Uji-shi, Kyoto 611, Japan*

Received 26 June; accepted 10 November 1978.

1. Nomura, N., Yamagishi, H. & Oka, A. *Gene* **3**, 39–52 (1978).
2. Nomura, N., Oka, A., Takanami, M. & Yamagishi, H. in *Single-stranded DNA Phages* (eds Denhardt, D. T., Dressler, D. H. & Ray, D. S.) (Cold Spring Harbor Laboratory, New York, in the press).
3. Davies, J. *et al.* in *R factors, their Properties and Possible Control* (eds Drew, J. & Högenauer, G.) 101–114 (Springer, New York, 1977).
4. Kopecko, D. J., Brevet, J. & Cohen, S. N. *J. molec. Biol.* **108**, 333–360 (1976).
5. Oka, A. *J. Bact.* **133**, 916–924 (1978).
6. Tomizawa, J. *Nature* **257**, 253–254 (1975).
7. Takanami, M. *FEBS Lett.* **34**, 318–322 (1973).
8. Maxam, A. M. & Gilbert, W. *Proc. natn. Acad. Sci. U.S.A.* **74**, 560–564 (1977).
9. Caros, M. P., Johnsrud, L. & Miller, J. H. *Cell* **13**, 411–418 (1978).
10. Grindley, N. D. F. *Cell* **13**, 419–426 (1978).
11. Musso, R. E. & Rosenberg, M. in *DNA Insertion Elements, Plasmids, and Episomes* (eds Bukhari, A. I., Shapiro, J. A. & Adhya, S. L.) 597–600 (Cold Spring Harbor Laboratory, New York, 1977).
12. Ohtsubo, H. & Ohtsubo, E. *Proc. natn. Acad. Sci. U.S.A.* **75**, 615–619 (1978).

The chromophore binding space of opsin

THE rate of the reaction between opsin and 11-*cis* retinal, that gives the visual pigment rhodopsin, is measured in the presence and absence of structural analogues of 11-*cis* retinal, the polyene side chain of which is varied systematically. A decrease in regeneration rate can be interpreted as competition with 11-*cis* retinal for the binding site on the protein. This approach seems to provide information on the molecular dimensions of the

chromophore binding space of opsin. Rotmans noticed that all-*trans* retinal does not inhibit the reaction between opsin and 11-*cis* retinal leading to rhodopsin¹. More recently Mastumoto and Yoshizawa² have reported that β -ionone, a compound carrying the six-membered ring configuration of retinal but with a shorter side chain, does inhibit the rate of this reaction. The surprising discrepancy between these two observations has led us to study more systematically the effects on the regeneration rate of a number of 11-*cis* retinal analogues, in which either the polyene side chain or the aldehyde group has been modified. We report here results that suggest that opsin possesses a specific chromophore binding space, one dimension of which is determined by the longest axis of 11-*cis* retinal, measured from the hydrophobic binding site onwards.

Opsin, virtually free of rhodopsin, retinal and retinol, was prepared by sucrose density gradient centrifugation of bovine retina homogenate illuminated in the presence of NADPH (ref. 3). Aliquots were stored at -70°C until used. 11-*cis* Retinal was isolated from a photoisomerate of retinal⁴ (Eastman) and 11-*cis* retinol was prepared by NaBH_4 reduction of 11-*cis* retinal⁵. The purity of all compounds used was checked by spectrophotometry and thin-layer chromatography.

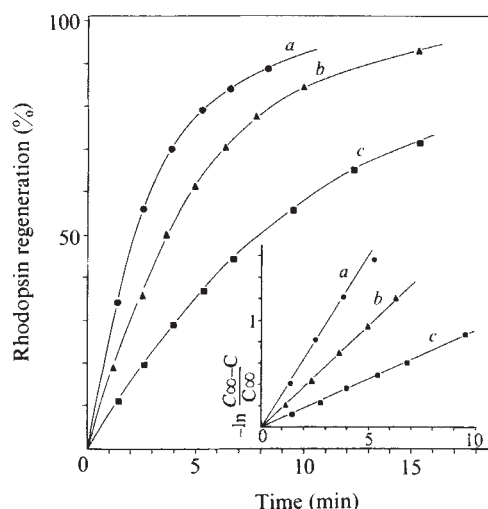


Fig. 1 Rate of the reaction between opsin and 11-*cis* retinal. *a*, In the absence of analogue; *b*, in the presence of the analogue 'C₁₈-ketone' (see Fig. 2*c*); *c*, in the presence of 'C₁₅-aldehyde' (see Fig. 2*b*). Reaction conditions, [opsin] = 9.5 μM ; [11-*cis* retinal] = 48.5 μM ; [analogue] = 500 μM ; pH 6.5; 20 $^{\circ}\text{C}$. Inset, semi-logarithmic plot.

The rate of the regeneration reaction was measured in a capped cuvette with two identical compartments. One compartment contained 600 μl of the opsin suspension (19 μM in phosphate buffer) to which 10 μl of ethanol was added, either with or without the short chain analogue to be tested. The other compartment contained an equal volume of a suspension of 11-*cis* retinal (97 μM in phosphate buffer, freshly prepared by adding the 11-*cis* retinal dissolved in 10 μl of ethanol). An initial absorption spectrum was recorded (Pye Unicam SP-1750, 20 $^{\circ}\text{C}$), and the reaction started by mixing the contents of the two compartments. At various time intervals the absorption spectrum was recorded between 650 and 450 nm. The increase in absorbance at 520 nm (corrected for scattering differences at 650 nm) is used to measure the rate of rhodopsin synthesis⁶. Alternatively, aliquots were taken from the reaction mixture at various time intervals. The regeneration reaction was stopped by addition of NH_2OH (final concentration 100 mM), followed after 10 min by Triton X-100 (final concentration 1%); 500 nm absorbance was then measured before and after illumination. Both methods give essentially the same results. Difference spectra always indicate the formation of rhodopsin with λ_{max} 500 nm.

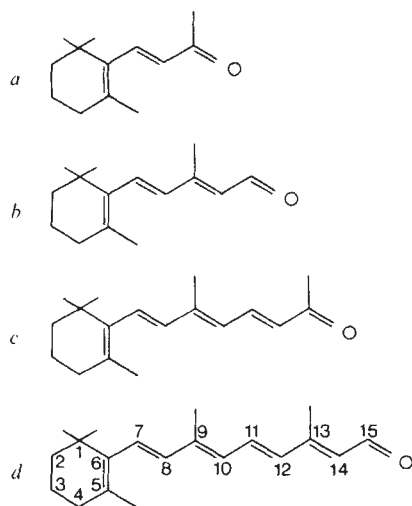


Fig. 2 Schematic structural formulae of short chain analogues of retinal. *a*, β -ionone; *b*, 'C₁₅-aldehyde', C₁₅H₂₂O, β -(3-methyl-5-oxo-1,3-all-*trans*-pentadienyl)2,6,6-trimethylcyclohexene; *c*, 'C₁₈-ketone', C₁₈H₂₆O, β -(3,7-dimethyl-7-oxo-1,3,5-all-*trans*-heptatrienyl)2,6,6-trimethylcyclohexene; *d* all-*trans* retinal.

Outer segment membrane suspensions rather than detergent solutions of opsin^{2,6,7} have been used to approach the natural microenvironment of the protein as closely as possible. A fivefold molar excess of 11-*cis* retinal with respect to opsin was used. This results, in our conditions, in almost perfect first order kinetics over a large part (>65%) of the regeneration reaction, both in the presence and absence of analogues (Fig. 1). When 11-*cis* retinal is added to opsin as a very concentrated solution, a biphasic reaction is observed with a very rapid first phase. The analogues were preincubated for 10 min in the opsin suspension in a 10-fold molar excess with respect to 11-*cis* retinal. In this way reproducible results can be obtained.

Table 1 Apparent pseudo first order rate constants of the reaction opsin + 11-*cis* retinal $\rightarrow$ rhodopsin, in the presence and absence of various retinal analogues

Analogue	$k_{app}(\text{min}^{-1})$
No addition	0.33
β -ionone	0.083
'C ₁₅ -aldehyde'	0.092
'C ₁₅ -alcohol'	0.084
'C ₁₈ -ketone'	0.20
All- <i>trans</i> retinol	0.19
11- <i>cis</i> retinol	0.18
All- <i>trans</i> retinal	0.32
13- <i>cis</i> retinal	0.36

Measurements of the regeneration rate of rhodopsin are presented in Fig. 1*a*. From the semilogarithmic plot (inset, *a*) a pseudo first order rate constant (k_{app}) of 0.33 min⁻¹ was calculated. With regard to the effects of the analogues on the reaction rate, three patterns can be distinguished (Fig. 1, Table 1): (1) no effect is seen with all-*trans* as well as with 13-*cis* retinal (Fig. 1*a*), (2) moderate inhibition is observed with the 'C₁₈-ketone', all-*trans* retinol and 11-*cis* retinol (Fig. 1*b*), (3) strong inhibition is caused by β -ionone, the 'C₁₅-aldehyde' and the 'C₁₅-alcohol' (Fig. 1*c*). Varying the concentration of inhibitors between 0.25 and 5 mM shows that competitive inhibition occurs; the presence of even a large excess of analogues never prevents complete regeneration of rhodopsin.

An explanation of these observations originates from the conclusion of Matsumoto and Yoshizawa², that β -ionone competes with 11-*cis* retinal for a binding site on opsin. We observe that the 'C₁₅-aldehyde' inhibits the regeneration reaction to the same extent. As in the case of β -ionone², the presence

of the aldehyde group is not essential, for the 'C₁₅-alcohol' gives essentially the same result. This seems to confirm the concept of a specific hydrophobic binding site for the trimethylcyclohexene ring on opsin. On extending the side chain of the analogues, the importance of a second parameter becomes apparent. The inhibition decreases markedly in the case of the 'C₁₈-ketone' and all-*trans* retinol. These two compounds contain a side chain with a rigid part consisting of four conjugated all-*trans* double bonds. It appears that the presence of a side chain, with a rigid part longer than that in the 11-*cis* retinal molecule, hinders the interaction of the six-membered ring moieties of these analogues with the hydrophobic binding site on opsin. This assumption seems to be supported by the fact that all-*trans* retinal, containing five conjugated all-*trans* double bonds in the side chain, does not inhibit the regeneration rate at all. Whereas the fairly high selectivity of retinal analogues with regard to the formation of visual pigments⁸⁻¹¹ suggests the existence of a rather specific chromophore binding space in rhodopsin, our findings indicate that such a specificity, with regard to the longest axis of the chromophore, is already present in the opsin molecule before the formation of visual pigment has taken place¹².

The effects of 11-*cis* retinal (moderate inhibition) and 13-*cis* retinal (no inhibition) are qualitatively compatible with our model. They may provide more refined information about the chromophore binding space of opsin, once their three-dimensional structure is known exactly.

In conclusion, competitive effects of retinal analogues with 11-*cis* retinal on the kinetics of the regeneration of rhodopsin from opsin provide information about the chromophore binding space of opsin. Apart from a previously noted² hydrophobic binding site for the six-membered ring of 11-*cis* retinal, it appears to be characterised by specific dimensions. The longest dimension of this space seems to coincide approximately with the long axis of the 11-*cis* retinal molecule.

I thank Cor Jacobs and Dr J. L. Baas for analogues. This investigation was carried out with financial support from the Netherlands Organization for Basic Research (ZWO) through the Netherlands Foundation for Chemical Research (SON).

F. J. M. DAEMEN

Department of Biochemistry,
University of Nijmegen,
Nijmegen, The Netherlands

Received 17 July; accepted 13 November 1978.

1. Rotmans, J. P. thesis, Univ. Nijmegen 89-91 (1973).
2. Matsumoto, H. & Yoshizawa, T. *Nature* **258**, 523-526 (1975).
3. de Grip, W. J., Daemen, F. J. M. & Bonting, S. L. *Vision Res.* **12**, 1697-1707 (1972).
4. Brown, P. K. & Wald, G. *J. biol. Chem.* **222**, 865-877 (1956).
5. Hubbard, R., Brown, P. K. & Bownds, D. in *Meth. Enzym.* **18**, 615-653 (1971).
6. Matsumoto, H., Horiuchi, K. & Yoshizawa, T. *Biochim. Biophys. Acta* **501**, 257-268 (1978).
7. Wald, G. & Brown, P. K. *Nature* **177**, 174-176 (1956).
8. Kropf, A., Whittenberger, B. P., Goff, S. P. & Waggoner, A. S. *Expl Eye Res.* **17**, 591-606 (1973).
9. Honig, B. & Ebrey, T. A. *Rev. Biophys. Bioengng* **3**, 151-177 (1974).
10. de Grip, W. J., Liu, R. S. H., Ramamurthy, V. & Asato, A. *Nature* **262**, 416-418 (1976).
11. Matsumoto, H. & Yoshizawa, T. *Vision Res.* **18**, 607-609 (1978).
12. Daemen, F. J. M. & Bonting, S. L. *Biophys. Struct. Mech.* **3**, 117-120 (1977).

Nature Index and Binders

The complete **Index** for 1977 is available, price £2.50 (UK), US\$5.00 (Rest of World). Copies of the 1976 index are still on sale, prices as above.

Binders for the journal are also available at £8.00 (UK), US\$16.00 (Rest of World) for three: a year of *Nature* fits into three binders.

All prices include postage. Payment may be made in any currency at the prevailing exchange rate. Orders should be sent, accompanied by remittance, to Macmillan Journals Ltd, Brunel Road, Basingstoke, Hampshire, RG21 2XS, UK.

matters arising

Premelting near crystal defects

BARTIS¹ has suggested that the premelting increment in heat capacity, which has been observed for some materials to be of the form

$$\Delta C = A/(T_m - T)^2 \quad (1)$$

may arise from melting below the intrinsic bulk melting point, in the vicinity of point defects. He also suggested that premature melting may occur in the vicinity of edge dislocations resulting in an incremental heat capacity of the form

$$\Delta C = B/(T_m - T)^3 \quad (2)$$

The important question arises as to whether the constants A and B are large enough to account for the observed magnitude of the premelting heat capacity. We show that B is too small for the dislocation contribution to be significant, but that for lattice vacancies A is about as large as the observed magnitude.

Gronvold² records that 99.99% purity bismuth is 2% molten 3 K below the atmospheric pressure melting point. By using standard expressions for the excess pressure about a dislocation³ one finds that a dislocation density in excess of 4×10^{14} lines m^{-2} is required to admit such a degree of premelting. Such a density is three to four orders of magnitude higher than the largest density one might expect for a stable dislocation array just below the melting point. For a more likely dislocation density at the melting point of 10^{10} lines m^{-2} bismuth would be 0.1% molten 0.07 K below its normal melting temperature which may be just within the limits of observation, but this may require a prohibitively high purity sample².

Neither, it seems, does the dislocation core melt at all. The problem of core melting of dislocations may be treated in the same manner as core sublimation as detailed by Frank⁴. He concluded that for all dislocations with Burgers' vector < 1 nm the core is not hollow. This includes all simple metals and ionic crystals. Similarly, in the case of melting below the bulk melting temperature, it can be shown⁵ that the melt radius for these crystals is always less than the core radius and therefore that core melting does not occur.

These ideas are borne out by experiment. Allnatt and Sime⁶ have studied the premelting rise in the electrical conductivity of NaCl after 5% compressional deformation and concluded that no contribution to the premelting arises from

the dislocations, and although the premelting surface conductivity is altered by deformation,⁷ this is not due to dislocation core melting. Moreover, Amelinckx and Dekeyser⁸, in reviewing the relevant experimental data, conclude that intrinsic melting does not measurably occur at grain boundaries.

It seems that premature melting near dislocations either does not occur or is so small as to be immeasurable. There may, however, be a possible source of premelting behaviour as described by equation (1), in the vicinity of vacancies. A vacancy in a metal usually results in a local decrease in pressure which, in the referenced case of bismuth, increases the local melting point, and this precludes vacancy-originated premelting. However, in the case of metals with normal melting curves the local melting temperature is depressed. A calculation of the magnitude of A in equation (1) reveals that the effect is sufficiently large to explain premelting over a degree or so below the melting point. This is illustrated by considering the depression of melting point that would occur if the total free energy of a system of vacancies were distributed uniformly through the crystal lattice. For a mole fraction α of vacancies with a Gibbs' function of formation of g_v each, in a crystal with a molar entropy of melting of ΔS_m the depression is approximately

$$\Delta T = \alpha N_A g_v / S_m, \quad (3)$$

where N_A is Avogadro's number. For the following typical values for metals: $g_v \sim 1$ eV, $\Delta S_m \sim 8.5$ J mol^{-1} K^{-1} and α ($T = T_m$) $\sim 10^{-4}$ we find $\Delta T \sim 1$ K. There may therefore be observable premature melting in the vicinity of lattice vacancies.

J. L. TALLON

Materials Science Section,
Physics and Engineering Laboratory,
Department of Scientific
and Industrial Research,
Lower Hutt, New Zealand

Present address: Department of Structural Properties of Materials, The Technical University of Denmark, DK 2800 Lyngby, Denmark.

1. Bartis, F. J. *Nature* **268**, 427–428 (1977).
2. Gronvold, F. *Acta chem. scand.* **A29**, 945–955 (1975).
3. Hirth, J. P. & Lothe, J. *Theory of Dislocations* 74 (McGraw-Hill, New York, 1968).
4. Frank, F. C. *Acta crystallogr.* **4**, 497–501 (1951).
5. Tallon, J. L. thesis Victoria Univ., New Zealand (1976).
6. Allnatt, A. R. & Sime, S. J. *Trans. Faraday Soc.* **67**, 674–678 (1971).
7. Tallon, J. L., Robinson, W. H. & Smedley, S. I. *J. phys. Chem.* **82**, 1277 (1978).
8. Amelinckx, S. & Dekeyser, W. in *Solid State Physics* (eds Seitz, F. & Turnbull, D.) **8**, 473 (1959).

BARTIS REPLIES—I question Tallon's assessment of the stresses around edge dislocations. If the stresses are weak, then

why are nucleation and growth alongside dislocations observed so often in structural studies of polymorphic transformations? Take, for instance, the passage of manganous oxide from the paramagnetic to the antiferromagnetic state. In neutron scattering from a single crystal above the Néel point Renninger *et al.*¹ detect peaks with satellites near the superlattice positions. They ascribe these features of the diffraction pattern to small domains of the magnetically ordered form of MnO. Using an electron microscope, Barber and Evans² do find small antiferromagnetic regions near dislocations and inclusions of Mn_3O_4 some 25 K above T_N . In the light of these results it is especially difficult to accept Tallon's claim that premelting does not occur in the cores of crystal dislocations.

F. J. BARTIS

Department of Physics,
Indiana University,
Bloomington, Indiana 47401

1. Renninger, A., Moss, S. C. & Averbach, B. L. *Phys. Rev.* **147**, 418–422 (1966).
2. Barber, D. J. & Evans, R. G. *J. mater. Sci.* **6**, 1237–1245 (1971).

The Golgi complex and inter-relationships of arthropods

LOCKE AND HUIE have reported¹ that the Golgi complex beads in a wide range of arthropods, but not in the Onychophora or other soft-bodied invertebrates, have a common feature—they stain with bismuth. They suggest that this favours a monophyletic rather than a polyphyletic concept of arthropod evolution. It should be remembered, however, that the taxa listed with similar Golgi staining also depend on unstriated muscle (which facilitates extensive shape changes of the body), in contrast to the sclerite-bearing arthropods. The significance of the similarity between Golgi phenomena and unstriated muscle is obscure. However, it is well known that striated muscle would not suit the requirements of the Onychophora^{2–6}. Also, it is certain that striated muscle has evolved more than once in the animal kingdom. The presence or absence of striated muscle in the Onychophora and in armoured arthropods seems to be of no more than functional significance.

Locke and Huie's opening paragraph does not seem to agree with present-day views on arthropod phylogeny. To take one point: by 1972 it was time to abandon the cumbersome phrase 'Onychophora—

Myriapoda-Hexapoda assemblage' and to give a name, the Uniramia, to this group⁷. The component taxa are united on their basic arthropodan features; their particular type of jaws, and their limbs lacking any trace of true outer ramus. The jointed limbs of myriapods and hexapods could, arguably, have been derived from lobopodial limbs essentially resembling those of the extant Onychophora, and these limbs could have been derived from limbs of metamERICALLY segmented worms with a haemocoel which, along with the muscles, worked the uniramous limbs. As long ago as 1940, Tiegs^{8,9} emphasised the unity of these three taxa and Anderson¹⁰ has shown in detail the common ontogenetic theme which exists throughout the Uniramia and contrasts absolutely with that pervading the entire crustacean group. Further, crustacean ontogeny could not have been derived from that of any known annelidan group. In contrast, the uniraman ontogeny might have been derived from that of some yolky-egged annelid or annelid-like worm which possessed a haemocoel and lobopodia, but not coelomate parapodia.

There is no justification on the basis of Golgi phenomena for the restriction of the term Uniramia to the Myriapoda and Hexapoda alone, nor for the exclusion from the arthropods of any animal on one character alone. That the Onychophora are thoroughly arthropodan is shown by their cuticle^{11,12}, which contains protein and chitin, but not collagen, in contrast to annelids; by the presence of an ecdysial cycle as in other arthropods¹³, the cuticle being shed in one piece, as in no other invertebrates on Locke and Huie's list¹; and by the general anatomy of the Onychophora, which tallies with that of other uniramian taxa. We do not know the significance of onychophoran Golgi phenomena nor of its association with unstriated muscle, but this one feature is no acceptable reason to disrupt the Uniramia or exclude the Onychophora from the arthropods.

Thus, without further consideration of the several other arthropodan taxa, it is clear that the Uniramia and Crustacea each represent well defined arthropodan taxa, which could not have originated from common annelidan ancestry, and that their cuticles must have evolved in parallel.

The exoskeleton of all arthropods has the same mechanical functions in providing unstretchable cuticles, sclerites and arthrodial membranes concerned with trunk movements and muscle insertions. The requirements of other invertebrates are different. The cuticles of the various arthropodan taxa are not identical, although protein and fine-fibre systems form the basis of most sclerites. There is thus no reason that the general manner of formation of arthropod cuticle may not have evolved more than once; nor is there any sound evidence against the unity of

the Uniramia or against a polyphyletic concept of arthropodan evolution.

S. M. MANTON

Zoology Department,
British Museum (Natural History),
London SW7, UK

1. Locke, M. & Huie, P. *Nature* **270**, 341-343 (1977).
2. Manton, S. M. *J. Linn. Soc., Zool.* **56**, 58-72 (1958).
3. Manton, S. M. *J. nat. Hist.* **1**, 1-22 (1967).
4. Manton, S. M. *J. Linn. Soc., Zool.* **53**, 257-375 (1973).
5. Manton, S. M. *J. zool. Res.* **171**, 111-130 (1974).
6. Manton, S. M. *The Arthropoda: Habits, Functional Morphology and Evolution* 1-527 (Oxford University Press 1977).
7. Manton, S. M. *J. Linn. Soc., Zool.* **51**, 203-400 (1972).
8. Tiegs, O. W. *Q. Jl. microsc. Sci.* **82**, 1-225 (1940); **85**, 191-328 (1945).
9. Tiegs, O. W. *Q. Jl. Microsc. Sci.* **88**, 165-336 (1947).
10. Anderson, D. T. *The Embryology and Phylogeny of Annelids and Arthropods* 1-495 (Pergamon, Oxford, 1973).
11. Robson, E. A. *Q. Jl. microsc. Sci.* **105**, 281-299 (1964).
12. Hackman, R. H. & Goldberg, M. *Science* **190**, 582-583 (1975).
13. Manton, S. M. *Phil. Trans. R. Soc. B.* **227**, 411-464.

Are viroids negative-strand viruses?

IN their paper detailing the complete sequence of 359 bases in potato spindle tuber viroid (PSTV) RNA, Gross *et al.*¹ considered, as have other workers², that the circular single-stranded molecule is unlikely to function as a messenger RNA. The possibility that PSTV RNA might be a non-coding strand rather than a positive (coding) strand may have been overlooked.

I have examined the base sequence complementary to that established by Gross *et al.*¹. Four possible initiation triplets (all GUG) and six possible strong termination triplets could result in four polypeptides containing 108, 79, 43 or 28 amino acids, of which the two larger (with molecular weights (MWs) 11,300 and 8,500) are the more interesting. The longest of these sequences begins with a GUG at positions 134-136 and terminates at a double UAG at positions 99-104 (Fig. 1). The GUG is preceded by a well-placed potential ribosome recognition sequence as indicated in Fig. 1. The second potential polypeptide is in a different reading frame and begins with a GUG at residues 177-179. It is preceded by an unbroken sequence of 14 purines, and terminates in an UAG at residues 55-57. The amino acid composition of the small polypeptide is unremarkable except for a high content of phenylalanine (10%), but the larger polypeptide has a content of arginine (11%) and lysine (5%) that would give it histone-like properties.

Two experimental observations would fit with the view that one or both of these polypeptides are produced by viroids. First, there is evidence for an RNA strand complementary to citrus exocortis viroid (CEV) RNA in infected tissues³; CEV is fairly closely related to PSTV². Some of the hybridisable RNA was in the nuclear fraction but more was in the soluble cytoplasmic fraction, where it might have possessed a messenger function.

Second, an increase in amount (and in radioactive labelling) of two polypeptide bands in polyacrylamide gels was found in CEV-infected plants⁴. MWs assigned to the two polypeptides were 18,000 and 15,000. I have recalculated these MWs as 13,200 and 10,400, using as internal markers the large and small subunits of ribulose-1',5'-diphosphatecarboxylase (MWs 54,000 and 12,000). The larger of the CEV-induced polypeptides was located preferentially in the histone fraction, the smaller was in the non-histone fraction⁴. This partitioning fits with the amino acid composition of the two polypeptide components that could be derived from the negative strand.

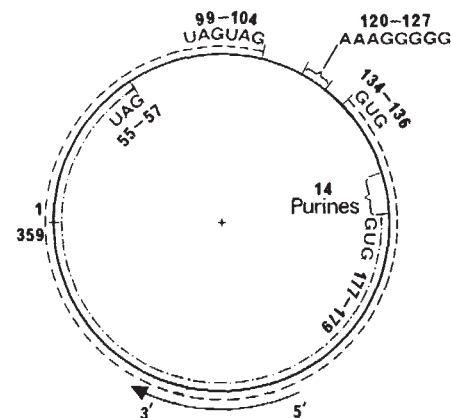


Fig. 1 Model of a PSTV RNA strand complementary to the viroid strand sequenced by Gross *et al.*¹. The heavy line represents the circular RNA. The numbers 1 and 359 indicate the same start position as that used by Gross *et al.*, but because the strand represented here is antiparallel to the viroid strand, the position of base 1 in this diagram is equivalent to that of base 359 in the viroid sequence. Dashed line, polypeptide of MW 11,300; dotted-dashed line, polypeptide of MW 8,500.

In the light of the above evidence I suggest that viroids may be very small negative-strand viruses with the following properties. (1) They use secondary structure rather than a protein coat for protection of the genome; (2) they use a host polymerase to make a complementary RNA strand; and (3) this strand serves as a template for progeny viroid particles and as message for one or two viroid proteins. This hypothesis can be tested by a search in PSTV-infected tissues for one or two viroid-induced polypeptides of the predicted size and composition.

R. E. F. MATTHEWS

Department of Cell Biology,
University of Auckland,
Private Bag, Auckland,
New Zealand

1. Gross, H. J. *et al.* *Nature* **273**, 203-208 (1978).
2. Dicner, T. O. & Hadidi, A. *Comprehensive Virol.* **11**, 285-337 (1977).
3. Grill, L. K. & Semancik, J. S. *Proc. natn. Acad. Sci. U.S.A.* **75**, 896-900 (1978).
4. Conejero, V. & Semancik, J. S. *Virology* **77**, 221-232 (1977).

reviews

Thoughts on ageing

A. J. S. Davies

Endurance of Life: Implications of Genetics for Human Life. By Macfarlane Burnet. Pp. 230. (Cambridge University Press: Cambridge and London, 1978.) £8.95.

PERHAPS the best of this new book of Burnet's is the information it gives about the man himself and the relationship he has to the society in which he has become famous. Burnet states that one of the reasons that he writes is to try to clarify his own mind. In this instance the supposed topic for elucidation is ageing—not surprisingly somewhat of an obsession at the age of 78. In fact the book is a collection of essays, of variable quality, loosely related by their common authorship but little else. Aside from the need for self-expression Burnet also senses that as an eminent man his book may well be read by others, some of whom will be laymen. His concessions to the non-scientist are, however, by no means uniform. Diagrams of childlike simplicity are followed by pages of relatively unexplained and sometimes unrelated complexities, and I believe it highly unlikely that any but the most dedicated non-scientist will persevere with the book—which is a pity. Science is greatly in need of expositors of its meaning and basic facts.

Burnet's approach to human affairs is essentially that of a rationalist. He believes that there are facts of life which are not widely recognised, perhaps because of their unpalatability, but which nevertheless should be given wider public airing. As a geneticist Burnet of course espouses the basic assumption that that which is perceived of an organism, its phenotype, arises as a consequence of an interaction between the genetic information encoded in the DNA, the genotype, and the environment. He is unimpressed by those who suppose that whereas this is all right for sweet peas and field-mice it is a gross over-simplification for *Homo sapiens*. Burnet's stance in this old nature versus nurture argument is resolute. He believes that such human characteristics as intelligence and aggression must have a strong genetic component, despite the fervent hopes of educationalists and politicians alike that this is not so.

Contemporary society, according to



Sir Macfarlane Burnet

Burnet, has a residue of activities, largely genetically determined, which were of advantage to the human species in the hunter-gatherer eras but for which the role has become less clear with the urbanisation and industrialisation of the past 10,000 years or so. A variety of social problems such as juvenile delinquency and the ennui of high rise flats (apartment blocks) seem to fall into perspective against such a back drop.

Statements of determinism have of course been made before, notably by Darlington. In Burnet they have engendered a massive pessimism arising perhaps from impotence at not knowing what to suggest if they are valid. Whatever the truth, the argument is a good one and is presented with a force which belies the years of the author.

Not content with a viewpoint which will antagonise almost every sociologist in the world, Burnet also points to the genetical homogenisation which he believes to be occurring in the middle echelons of society which could act to diminish human diversity to an undesirable degree. Reduction in family size he notes may prevent overpopulation but it will also disturb a natural balance which involved much natural

selection within larger families. Burnet has also taken to heart the argument that one of the less desirable outcomes of modern medicine is the increasing preservation of the genetically unfit.

Burnet stops more or less short of the conclusions of the pre-war eugenicists but he nevertheless feels strongly that genetic counselling is worthwhile. He also believes severely handicapped children might well not be given the life support that they require.

He is an elitist who sees harm in differential breeding rates in the community which sometimes seem to favour the increase in the size of the populations of the under-privileged, but again he has no specific recommendations to make. His arguments are not in the main supported by an adequate bibliographical background but their emotional content is clear. Burnet sees himself as speaking with a common sense voice for the majority of intelligent men and women.

There are in the book many paradoxes. Burnet seems to equate power with evil on the basis of a definition of power as influence over people but he fails to realise that he himself is an extremely powerful and influential person but probably not evil. Goodness seems to be largely a feminine characteristic, a viewpoint which ought to make the more militant female liberationists fume.

The least satisfactory of the essays are those concerned with ageing. Briefly, Burnet sees ageing as due to the accumulation of genetic errors within somatic cells. Particularly important are errors in the error-rectifying system, which should have far more disastrous consequences than more mundane coding alterations. Burnet accepts all evidence for his theory but often seems disinterested in what scientists call the null hypothesis. He even has the face to criticise the virologists (p 39) for apparently adopting a similar attitude. Indeed Burnet's hostility to viruses borders on the irrational in that he seems unable to envisage that one of the genetic errors in somatic cells which he invokes so freely might in fact be the incorporation of a viral genome.

There is a whole chapter on *Xeroderma pigmentosum* which is seen, perhaps correctly, as a condition in

which a deficient error repair mechanism leads to a variety of defects including a high susceptibility to ultra-violet light and increasing numbers of skin tumours. Burnet sees this unpleasant syndrome, which has a known genetic basis, as a grossly accelerated example of ageing corresponding exactly to his hypothesis. In addition the belief he has about cancers arising as a consequence of somatic mutations is confirmed. To one who has read a number of Burnet's previous books much of the early parts of the present volume may evoke a strong sense of *déjà vu*.

As, with succeeding chapters, Burnet moves from Ageing in Mammals by way of the Genetics of Power to his Vision of a Million Years the book becomes more readable. Those who wish to dip into it should in my opinion start at the back and work forward until they come to familiar ground. The most evocative and fluent passages are those which concern Burnet himself. He talks wistfully of societies in which privileged elders can retain an honoured position as men of experience and transmitters of traditional wisdom. Despite or perhaps because of this he makes statements of astonishing lack of insight—for example, "a little thought will con-

vince most people that natural death means a death in which the essential factors concerned were genetic ones". The value judgment involved here in the use of the adjective essential and the apparent lack of realisation of the operation of the many stochastic events which could influence the matter are surprising.

Whenever he himself has not thought about a particular matter it is dismissed in a lordly manner—for example, "we need not worry too hard to imagine by what mechanism it ('psychosomatic' death may be achieved."

Atomic and molecular physics

Atoms and Molecules. By M. Weissbluth. Pp.713. (Academic: New York, San Francisco and London, 1978.) \$59.50; £38.65.

GROWN out of a full year graduate lecture course in atomic and molecular physics, this monograph is both a textbook and a reference manual of ambitiously wide scope. Thus the final chapter (28) on ligand fields treats such topics as electronic d^2 in a cubic field with spin-orbit coupling. And the chapter's last section concludes with a diagram that shows the energy levels of molecular orbitals when an octahedral complex AB_6 is involved; from a preceding table follows, for each symmetry associated with the levels, the structure of symmetry adapted orbitals for both the orbitals of the central atom A and the ligands B. It would be only a short step to frontier problems such as ferric ions in proteins. Derivations often take just a few lines—try this conclusively without simultaneously applying group theory to the central electrons and the lattice-like structure around them! All the required atomic and algebraic tools and data are lucidly expanded in the book, and referencing to them is done efficiently.

Naturally group theoretical topics take up most of the space in the first of the text's six parts, which gives the mathematical background. Only rigorous pruning to the needs of atomic and molecular physics, and resorting to summaries of selected properties where appropriate, keeps this part in size and readable. He who struggled through the thicket of the full representation theory of the permutation group, this centrepiece in problems involving identical particles, will appreciate the refreshing brevity and swiftness with which the relevant steps in Young's scheme evolve.

Part II, on the quantum-mechanical background, assumes some undergraduate course knowledge. The remaining four parts move on towards molecules: one-electron atoms (from Dirac's equation

Few aspects of our contemporary society escape Burnet's pessimistic pen. Most of his views will be unpopular and one cannot help but agree with his statement that he has never failed to write, or modify what he has written, out of fear of the consequences. The book is erratic, aggravating, pontifical on occasions, sometimes clumsy, often obscure, but rarely dull. It should be read and discussed. □

A. J. S. Davies is Head of the Biology Division at the Chester Beatty Research Institute, London, UK.

by way of Schrödinger's to hyperfine interactions), N-electron atoms, electromagnetic interactions (culminating in the Kramers-Heisenberg formula and in applications), and finally 130 pages on molecules, covering in particular the spectra of polyatomic molecules. The notion of conjugate configuration, in the context of N-electron problems, tells immediately why no spin-orbit splitting occurs for half-filled shells. However, even then the mutual effect between electrons does give rise to some splitting. Such splitting may be ignored because of its smallness, though the author, in this rare instance, misses the cause when relating it to interaction solely between terms. All except purists are likely to overlook an unnecessary impediment to more fundamental approaches in some basic relationships: to equivalence a unit vector directly with its cartesian representation as a column matrix, and on occasion a state vector with the complete function—something quite commonly found in textbooks. Despite these shortcomings the whole presentation is elegant and informative.

This new book shares one weakness with many a first edition: the index doesn't measure up to the text. Similarly, the references haven't quite outgrown a list for the students of recommended reading. There is one particular omission: among the three classics on group theory and quantum mechanics, which came out within a very short period around 1931, H. Weyl's and E. Wigner's books are mentioned, whereas B. L. van der Waerden's is not.

Who is going to buy the book? Sadly enough the price tag will turn away most graduate and postgraduate students. Needless to say that the book is well produced and the typographical layout excellent. But perhaps one should draw attention to over 70 useful tables, an impressive collection of information which, leaving aside the newly compiled tables, one might as well find scattered across several book-shelves.

Werner Eissner

Werner Eissner is Temporary Senior Lecturer in the Department of Physics and Astronomy at University College, London, UK.

BIO-MATHEMATICS Revue de Bio-Mathématique

Founded in 1962

Bilingual English, French

4 issues per year

If direct application of mathematics to biology interests you, if you think that concepts such as Form, Structure, Complexity of living organisms are worth putting into mathematical language, or that the concept of Entropy and Information may play an important part in the evolution of the living world, "Bio-Mathematics" will give you something to think about.

A list of published papers is available

Editions Européennes
11 bis Avenue de
la Providence
92160 Antony. France.

Subscription 65F